

GenCore version 5.1.6  
copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw modeL  
Run on: July 3, 2003, 14:05:35 ; Search time 17.5 Seconds

(without alignments) 98.565 Million cell updates/sec

Title: US-09-830-839-6  
Perfect score: 77  
Sequence: 1 WNFAGIEAAASAIQG 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues  
Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

```
1: /cggn2_6/ptodata/1/pubpaas/US08_NEW_PUB.pep:*
2: /cggn2_6/ptodata/1/pubpaas/PCTNS_PUBCOMB.pep:*
3: /cggn2_6/ptodata/1/pubpaas/US05_NEW_PUB.pep:*
4: /cggn2_6/ptodata/1/pubpaas/US06_PUBCOMB.pep:*
5: /cggn2_6/ptodata/1/pubpaas/US07_NEW_PUB.pep:*
6: /cggn2_6/ptodata/1/pubpaas/US07_PUBCOMB.pep:*
7: /cggn2_6/ptodata/1/pubpaas/PCTNS_PUBCOMB.pep:*
8: /cggn2_6/ptodata/1/pubpaas/US08_PUBCOMB.pep:*
9: /cggn2_6/ptodata/1/pubpaas/US09_NEW_PUB.pep:*
10: /cggn2_6/ptodata/1/pubpaas/US09_PUBCOMB.pep:*
11: /cggn2_6/ptodata/1/pubpaas/US10_NEW_PUB.pep:*
12: /cggn2_6/ptodata/1/pubpaas/US10_PUBCOMB.pep:*
13: /cggn2_5/ptodata/1/pubpaas/US09_NEW_PUB.pep:*
14: /cggn2_6/ptodata/1/pubpaas/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	77	100.0	20	9 US-10-044-703-60
2	77	100.0	20	10 US-09-813-333-60
3	99	100.0	95	9 US-09-805-427A-1
4	77	100.0	403	9 US-09-805-427A-4
5	77	100.0	403	10 US-09-791-171-173
6	77	100.0	404	9 US-09-805-427A-3
7	77	100.0	404	10 US-09-791-171-172
8	66	85.7	20	9 US-10-044-103-61
9	66	85.7	20	10 US-09-813-333-61
10	60	77.9	13	9 US-10-044-703-59
11	60	77.9	13	10 US-09-813-333-59
12	41	53.2	463	9 US-10-761-12653
13	39	50.6	9	10 US-09-916-201-13
14	39	50.6	221	9 US-10-054-988-155
15	39	50.6	221	10 US-09-739-615-155
16	39	50.6	221	10 US-09-904-615-155
17	39	50.6	239	9 US-10-103-313-351
18	38	49.4	37	10 US-09-761-39026
19	48.1	153	10	US-09-925-301-1367

### ALIGNMENTS

```
RESULT 1
US-10-044-703-60
; Sequence 60, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIORITY APPLICATION NUMBER: 60/190,834
; PRIORITY FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-044-703-60

Query Match 100.0%; Score 77; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-06; Mismatches 0; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNFAGIEAAASAIQG 15
Db 2 WNFAGIEAAASAIQG 16

; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIORITY APPLICATION NUMBER: 60/190,834
; PRIORITY FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
```

;

; SOFTWARE: Patentin ver. 2.1

; SEQ ID NO: 60

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

US-09-813-332-60

Query Match 100.0%; Score 77; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAASATOG 15  
Db 2 WNFAGIEAASATOG 16

RESULT 3

US-09-805-427A-1

Sequence 1, Application US/09805427A

Patent No. US2002176867A1

GENERAL INFORMATION:

APPLICANT: Statens Serum Institut

TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens

FILE REFERENCE: 670001-2002.5

CURRENT APPLICATION NUMBER: US/09/805,427A

CURRENT FILING DATE: 2001-03-13

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin version 3.0

SEQ ID NO 1

LENGTH: 95

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-09-805-427A-1

Query Match 100.0%; Score 77; DB 9; Length 95;  
Best Local Similarity 100.0%; Pred. No. 8.6e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAASATOG 15  
Db 6 WNFAGIEAASATOG 20

RESULT 4

US-09-805-427A-4

Sequence 4, Application US/09805427A

Patent No. US2002176867A1

GENERAL INFORMATION:

APPLICANT: Statens Serum Institut

TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens

FILE REFERENCE: 670001-2002.5

CURRENT APPLICATION NUMBER: US/09/805,427A

CURRENT FILING DATE: 2001-03-13

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin version 3.0

SEQ ID NO 4

LENGTH: 403

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag85B

US-09-805-427A-4

Query Match 100.0%; Score 77; DB 9; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAASATOG 15  
Db 27 WNFAGIEAASATOG 41

RESULT 5

US-09-791-171-173

Sequence 173, Application US/09791173

Patent No. US20020094336A1

GENERAL INFORMATION:

APPLICANT: ANDERSEN, Peter

APPLICANT: NIELSEN, Rikke

APPLICANT: OETTINGER, Thomas

APPLICANT: RASMUSSEN, Peter Birk

APPLICANT: ROSEKRAKS, Ida

APPLICANT: WELDINGH, Karin

APPLICANT: FLORIO, Walter

TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS DERIVED FROM M. TUBERCULOSIS

FILE REFERENCE: 670001-002.1

CURRENT APPLICATION NUMBER: US/09/791,171

CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: 09/050,739

PRIOR FILING DATE: 1998-03-30

PRIOR APPLICATION NUMBER: 03/16/97

PRIOR FILING DATE: 1997-04-02

PRIOR APPLICATION NUMBER: 1277/97

PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: 60/044,624

PRIOR FILING DATE: 1997-04-18

PRIOR APPLICATION NUMBER: 60/070,488

PRIOR FILING DATE: 1998-01-05

NUMBER OF SEQ ID NOS: 173

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 173

LENGTH: 403

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-09-791-171-173

Query Match 100.0%; Score 77; DB 10; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAASATOG 15  
Db 27 WNFAGIEAASATOG 41

RESULT 6

US-09-805-427A-3

Sequence 3, Application US/09805427A

Patent No. US2002176867A1

GENERAL INFORMATION:

APPLICANT: Statens Serum Institut

TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens

FILE REFERENCE: 670001-2002.5

CURRENT APPLICATION NUMBER: US/09/805,427A

CURRENT FILING DATE: 2001-03-13

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin version 3.0

SEQ ID NO 3

LENGTH: 404

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Fusion protein Ag85B-ESAT-6

US-09-805-427A-3

Query Match 100.0%; Score 77; DB 9; Length 404;  
Best Local Similarity 100.0%; Pred. No. 4e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAASATOG 15  
Db 315 WNFAGIEAASATOG 329

RESULT 7

US-09-791-171-172

; Sequence 172, Application US/09791171

; Patent No. US20020094336A1

; GENERAL INFORMATION:

; APPLICANT: Andersen, Peter

; APPLICANT: NIELSEN, Rikke

; APPLICANT: OETTINGER, Thomas

; APPLICANT: RASMUSSEN, Peter Birk

; APPLICANT: ROSENKRANDS, Ida

; APPLICANT: WELDINGH, Karin

; APPLICANT: FLORIO, Walter

; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS

; TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS

; FILE REFERENCE: 670001-2002.1

; CURRENT APPLICATION NUMBER: US/09/791,171

; CURRENT FILING DATE: 2001-02-20

; PRIOR APPLICATION NUMBER: 09/050,739

; PRIOR FILING DATE: 1998-03-30

; PRIOR APPLICATION NUMBER: 0376/97

; PRIOR FILING DATE: 1997-04-02

; PRIOR APPLICATION NUMBER: 1277/97

; PRIOR FILING DATE: 1997-11-10

; PRIOR APPLICATION NUMBER: 60/044,624

; PRIOR FILING DATE: 1997-04-18

; PRIOR APPLICATION NUMBER: 60/070,488

; PRIOR FILING DATE: 1998-01-05

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 172

; LENGTH: 404

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-791-171-172

; Query Match 100.0%; Score 77; DB 10; Length 404;

; Best Local Similarity 100.0%; Pred. No. 4e-05; Mismatches 0; Indels 0; Gaps 0;

; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Qy 1 WNFAGIEAAASAIQ 15

; Db 315 WNFAGIEAAASAIQ 329

; RESULT 8

; US-10-044-703-61

; Sequence 61, Application US/10044703

; Publication No. US20020192233A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/10/044,703

; CURRENT FILING DATE: 2002-05-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-10-044-703-59

; Query Match 85.7%; Score 66; DB 9; Length 20;

; Best Local Similarity 100.0%; Pred. No. 0.00012; Mismatches 0; Indels 0; Gaps 0;

; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Qy 2 NFAGIEAAASAIQ 14

; Db 1 NFAGIEAAASAIQ 13

; RESULT 9

; US-10-044-703-61

; ORGANISM: Mycobacterium tuberculosis

; Query Match 85.7%; Score 66; DB 9; Length 20;

; Best Local Similarity 100.0%; Pred. No. 0.00012; Mismatches 0; Indels 0; Gaps 0;

; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; Qy 2 NFAGIEAAASAIQ 15

; Db 1 NFAGIEAAASAIQ 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-61

; Sequence 61, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT: Degroot, Anne S

; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters

; FILE REFERENCE: 17999-004 US

; CURRENT APPLICATION NUMBER: US/09/813,333

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/190,834

; PRIOR FILING DATE: 2000-03-20

; NUMBER OF SEQ ID NOS: 81

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 59

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-813-333-59

; Sequence 59, Application US/09813333

; Patent No. US2002019160A1

; GENERAL INFORMATION:

; APPLICANT:

; TYPE: PRG  
; ORGANISM: Mycobacterium tuberculosis  
; US-09-813-333-59

Query Match 77.9%; Score 60; DB 10; Length 13;  
Best Local Similarity 100.0%; Pred. No. 0.0006; Length 13;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 NFAGIEAAASAIQ 14  
Db 1 NFAGIEAAASAIQ 13

RESULT 12

US-10-156-761-12653

;

Sequence 12653, Application US/10156761

;

GENERAL INFORMATION:

;

APPLICANT: OMURA, SATOSHI

;

APPLICANT: IKEDA, HARUO

;

APPLICANT: ISHIKAWA, JUN

;

APPLICANT: HORIKAWA, HIROSHI

;

APPLICANT: SHIBA, TADAYOSHI

;

APPLICANT: SAKAKI, YOSHITAKI

;

APPLICANT: HAYTORI, MASAHIRA

;

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

;

FILE REFERENCE: 249-262

;

CURRENT APPLICATION NUMBER: US/10/156,761

;

CURRENT FILING DATE: 2002-03-29

;

PRIOR APPLICATION NUMBER: JP 2001-204089

;

PRIOR FILING DATE: 2001-05-30

;

PRIOR APPLICATION NUMBER: JP 2001-272697

;

PRIOR FILING DATE: 2001-08-02

;

NUMBER OF SEQ ID NOS: 15109

;

SEQ ID NO 12653

;

LENGTH: 463

;

TYPE: PRG

;

ORGANISM: Streptomyces avermitillis

;

US-10-156-761-12653

Query Match 50.6%; Score 39; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.9e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GIERASAT 13  
Db 1 GIERASAT 9

RESULT 14

US-10-054-988-155

;

Sequence 155, Application US/10054988

;

Publication No. US20030087341A1

;

GENERAL INFORMATION:

;

APPLICANT: Rosen et al.

;

TITLE OF INVENTION: 49 Human Secreted Proteins

;

FILE REFERENCE: P2032P1

;

CURRENT APPLICATION NUMBER: US/10/054,988

;

CURRENT FILING DATE: 2002-01-25

;

PRIOR APPLICATION NUMBER: 09/904,615

;

PRIOR FILING DATE: 2001-07-16

;

PRIOR APPLICATION NUMBER: 09/511,554

;

PRIOR FILING DATE: 2000-02-23

;

PRIOR APPLICATION NUMBER: 60/097,917

;

PRIOR FILING DATE: 1998-08-25

;

PRIOR APPLICATION NUMBER: 60/098,634

;

PRIOR FILING DATE: 1998-08-31

;

NUMBER OF SEQ ID NOS: 170

;

SOFTWARE: PatentIn Ver. 2.0

;

SEQ ID NO 155

;

LENGTH: 221

;

TYPE: PRG

;

ORGANISM: Homo sapiens

;

US-09-739-254-155

Query Match 50.6%; Score 39; DB 10; Length 221;  
Best Local Similarity 57.1%; Pred. No. 53; Length 221;  
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 5 GIERASAT 13  
Db 1 GIERASAT 9

RESULT 15

US-09-739-254-155

;

Sequence 155, Application US/09739254

;

Patent No. US200301700A1

;

GENERAL INFORMATION:

;

APPLICANT: Rosen et al.

;

TITLE OF INVENTION: 49 Human Secreted Proteins

;

FILE REFERENCE: P2032P1

;

CURRENT APPLICATION NUMBER: US/09/739,254

;

CURRENT FILING DATE: 2000-12-19

;

EARLIER APPLICATION NUMBER: 09/511,554

;

EARLIER FILING DATE: 2000-02-23

;

EARLIER APPLICATION NUMBER: PCT/US99/19330

;

EARLIER FILING DATE: 1999-08-24

;

EARLIER APPLICATION NUMBER: 60/097,917

;

EARLIER FILING DATE: 1998-08-25

;

EARLIER APPLICATION NUMBER: 60/098,634

;

EARLIER FILING DATE: 1998-08-31

;

NUMBER OF SEQ ID NOS: 170

;

SOFTWARE: PatentIn Ver. 2.0

;

SEQ ID NO 155

;

LENGTH: 221

;

TYPE: PRG

;

ORGANISM: Homo sapiens

;

US-09-739-254-155

Query Match 50.6%; Score 39; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.9e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GIERASAT 13  
Db 1 GIERASAT 9

RESULT 13

US-09-916-201-13

;

Sequence 13, Application US/09916201

;

Patent No. US20020134976A1

;

GENERAL INFORMATION:

;

APPLICANT: LALVANI, Ajit

;

APPLICANT: PATHAN, Ansar A.

;

APPLICANT: HILL, Adrian V.S.

;

TITLE OF INVENTION: TUBERCULOSIS VACCINE

;

FILE REFERENCE: 11-359

;

CURRENT APPLICATION NUMBER: US/09/916,201

;

CURRENT FILING DATE: 2001-07-27

;

PRIOR APPLICATION NUMBER: US 09/467,893

;

PRIOR FILING DATE: 1999-12-21

;

PRIOR APPLICATION NUMBER: US 60/113,783

;

PRIOR FILING DATE: 1998-12-23

;

NUMBER OF SEQ ID NOS: 17

;

NUMBER OF SEQ ID NOS: 17

;

SOFTWARE: PatentIn version 3.0

;

SEQ ID NO 13

;

LENGTH: 9

;

TYPE: PRG

;

ORGANISM: M. tuberculosis

;

US-09-916-201-13

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

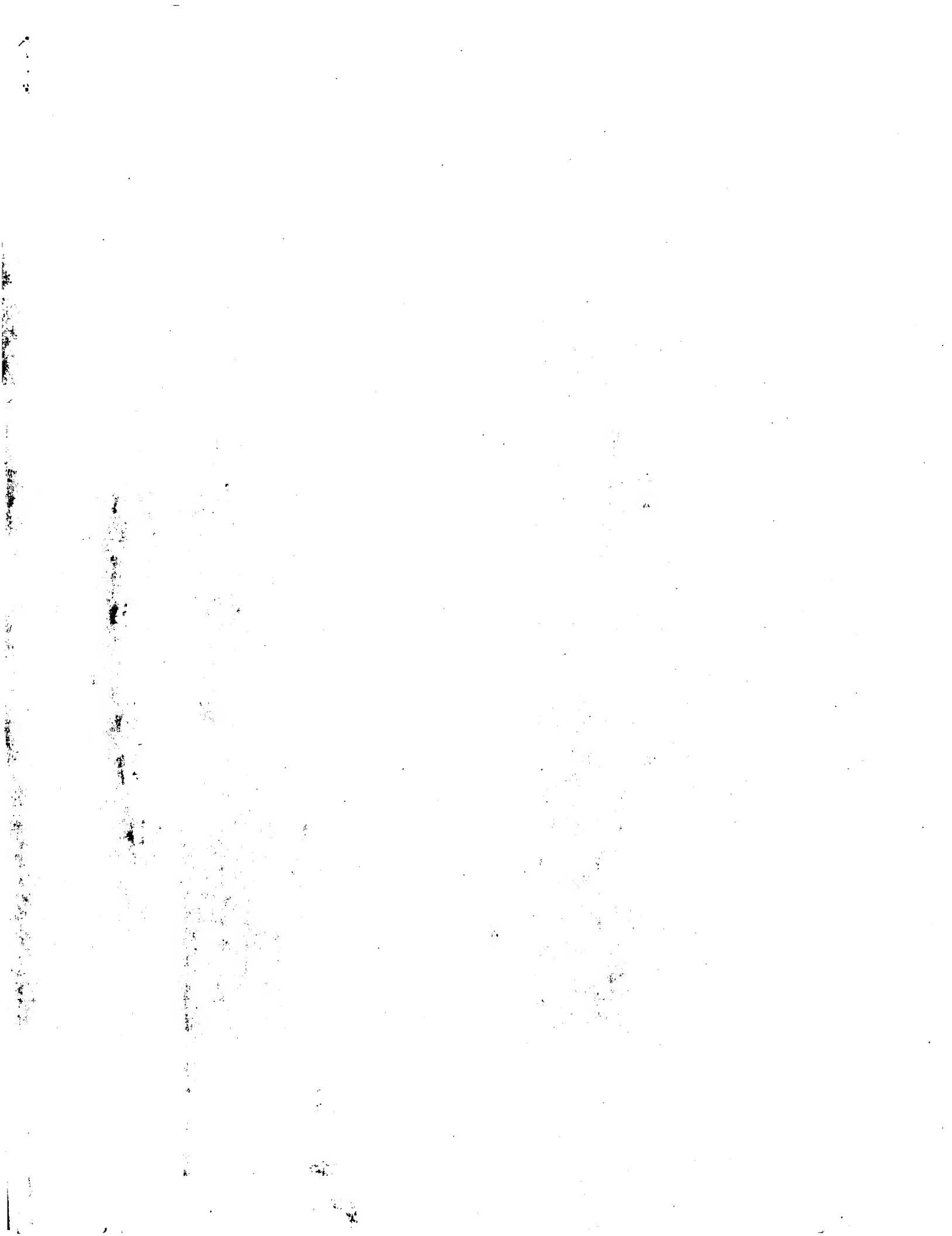
Thu Jul 3 14:50:03 2003

us-09-830-839-6.rapp

Page 5

OY 1 WNFAGTEAAASAIQ 14  
| | | | | | | | | |  
132 WNLGDEAAELAQ 145  
Db

Search completed: July 3, 2003, 14:08:27  
Job time : 17.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

On protein - protein search, using sw model

Run on: July 3, 2003, 14:04:30 ; Search time 11 Seconds  
(without alignments)  
40.122 Million cell updates/sec

Title:

Perfect score: US-09-830-839-6

Sequence: 1 WNFAGIEAAASATQG 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*

1: /cgn2\_6/prodata/1/iaa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*

3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/PCITS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	77	100	51	US-08-818-112-104
2	77	100	51	US-08-818-111-99
3	77	100	51	US-09-056-556-104
4	77	100	51	US-09-056-556-99
5	77	100	95	US-08-465-640-2
6	77	100	2293	US-09-368-590-2
7	38	49	418	US-09-202-838-4
8	37	48	1	US-09-509-902A-9
9	35	45	29	US-08-482-121-11
10	35	45	29	US-08-482-142-74
11	35	45	29	US-08-478-572-11
12	35	45	29	US-08-478-572-74
13	35	45	29	US-08-484-206-11
14	35	45	29	US-08-484-296-74
15	35	45	29	PCM-US05-04481-3
16	35	45	17	US-09-442-353-232
17	35	45	17	US-08-758-417A-80
18	35	45	159	US-09-447-803-18
19	35	45	222	US-07-945-208-11
20	35	45	222	US-08-462-801-11
21	35	45	222	US-08-461-809-11
22	35	45	222	US-08-461-441-11
23	35	45	222	PCT-US93-08518-11
24	35	45	245	US-07-945-208-2
25	35	45	245	Best Local Similarity
26	35	45	245	Matches 100.0%; Pred. No. 7.6e-07;
27	35	45	1	Mismatches 0; Indels 0; Gaps 0;
1	US-08-461-809-2	US-08-461-441-2	Sequence 2, Appli	

#### ALIGNMENTS

RESULT 1	US-08-818-112-104	; Sequence 104, Application US/08818112
		; Patent No. 6220969
		GENERAL INFORMATION:
		APPLICANT: Reed, Steven G.
		APPLICANT: Skeiky, Yasir A.W.
		APPLICANT: Dillon, Davin C.
		APPLICANT: Campos-Neto, Antonio
		APPLICANT: Houghton, Raymond
		APPLICANT: Vedwick, Thomas S.
		APPLICANT: Twardzik, Daniel R.
		TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
		NUMBER OF SEQUENCES: 153
		CORRESPONDENCE ADDRESS:
		ADDRESSEE: SPED and BERRY LLP
		STREET: 6330 Columbia Center, 701 Fifth Avenue
		CITY: Seattle
		STATE: Washington
		COUNTY: USA
		ZIP: 98104-7092
		COMPUTER READABLE FORM:
		MEDIUM TYPE: Floppy disk
		COMPUTER: IBM PC Compatible
		OPERATING SYSTEM: PC-DOS/MS-DOS
		SOFTWARE: Patentin Release #1.0, Version #1.30
		CURRENT APPLICATION DATA:
		APPLICATION NUMBER: US/08/818,112
		FILING DATE: 13-MAR-1997
		CLASSIFICATION: 424
		ATTORNEY/AGENT INFORMATION:
		NAME: Maki, David J.
		REGISTRATION NUMBER: 31,392
		REFERENCE/DOCKET NUMBER: 210121.411C6
		TELECOMMUNICATION INFORMATION:
		TELEPHONE: (206) 622-4500
		TELEFAX: (206) 682-6031
		INFORMATION FOR SEQ ID NO: 104:
		SEQUENCE CHARACTERISTICS:
		LENGTH: 51 amino acids
		TYPE: amino acid
		STRANDEDNESS: single
		TOPOLOGY: linear
		US-08-818-112-104

Query Match 100.0%; Score 77; DB 4; Length 51;  
Best Local Similarity 100.0%; Pred. No. 7.6e-07;  
Matches 100.0%; Mismatches 0; Indels 0; Gaps 0;

RESULT 2		US-08-818-111-99	
Sequence 99, Application US/08818111		STATE: Washington	
Patent No. 6338852		COUNTRY: USA	
GENERAL INFORMATION:		ZIP: 98104-7092	
APPLICANT: Reed, Steven G.		COMPUTER READABLE FORM:	
APPLICANT: Dillon, Davin C.		MEDIUM TYPE: Floppy disk	
APPLICANT: Campos-Neto, Antonia		OPERATING SYSTEM: PC-DOS/MS-DOS	
APPLICANT: Houghton, Raymond		SOFTWARE: PatentIn Release #1.0, Version #1.30	
APPLICANT: Vedvick, Thomas S.		CURRENT APPLICATION DATA:	
APPLICANT: Twardzik, Daniel R.		APPLICATION NUMBER: US/09/056,556	
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF		FILING DATE: 07-APR-1998	
NUMBER OF SEQUENCES: 148		CLASSIFICATION:	
CORRESPONDENCE ADDRESS:		ATTORNEY/AGENT INFORMATION:	
ADDRESSEE: SPEED and BERRY LLP		NAME: Maki, David J.	
STREET: 6300 Columbia Center, 701 Fifth Avenue		REGISTRATION NUMBER: 31,392	
CITY: Seattle		REFERENCE/DOCKET NUMBER: 210121.417C6	
STATE: Washington		TELECOMMUNICATION INFORMATION:	
COUNTRY: USA		TELEPHONE: (206) 682-4900	
ZIP: 98104-7092		TELEFAX: (206) 682-6031	
COMPUTER READABLE FORM:		INFORMATION FOR SEQ ID NO: 104:	
MEDIUM TYPE: Floppy disk		SEQUENCE CHARACTERISTICS:	
COMPUTER: IBM PC compatible		LENGTH: 51 amino acids	
OPERATING SYSTEM: PC-DOS/MS-DOS		TYPE: amino acid	
SOFTWARE: PatentIn Release #1.0, Version #1.30		STRANDEDNESS: single	
CURRENT APPLICATION DATA:		TOPOLOGY: linear	
APPLICATION NUMBER: US/08/818,111		US-09-056-556-104	
FILING DATE: 13-MAR-1997		Query Match 100.0% Score 77; DB 4; Length 51;	
CLASSIFICATION: 424		Best Local Similarity 100.0%; Pred. No. 7.6e-07;	
ATTORNEY/AGENT INFORMATION:		Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
NAME: Maki, David J.		QY 1 WNFAGIEAAASAIQG 15	
REGISTRATION NUMBER: 31,392		DB 6 WNFAGIEAAASAIQG 20	
REFERENCE/DOCKET NUMBER: 210121.417C6		RESULT 4	
TELECOMMUNICATION INFORMATION:		US-09-072-596-99	
TELEPHONE: (206) 682-4900		Sequence 99, application US/09072596	
TELEFAX: (206) 682-6031		Patent No. 6458366	
INFORMATION FOR SEQ ID NO: 99:		GENERAL INFORMATION:	
SEQUENCE CHARACTERISTICS:		APPLICANT: Reed, Steven G.	
LENGTH: 51 amino acids		APPLICANT: Dillon, Davin C.	
TYPE: amino acid		APPLICANT: Campos-Neto, Antonia	
STRANDEDNESS: single		APPLICANT: Houghton, Raymond	
TOPOLOGY: linear		APPLICANT: Vedvick, Thomas S.	
US-08-818-111-99		APPLICANT: Twardzik, Daniel R.	
Query Match 100.0% Score 77; DB 4; Length 51;		APPLICANT: Lodes, Michael J.	
Best Local Similarity 100.0%; Pred. No. 7.6e-07;		APPLICANT: Hendrickson, Ronald C.	
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF	
QY 1 WNFAGIEAAASAIQG 15		NUMBER OF SEQUENCES: 350	
DB 6 WNFAGIEAAASAIQG 20		CORRESPONDENCE ADDRESS:	
COMPUTER READABLE FORM:		ADDRESSEE: SPEED and BERRY LLP	
MEDIUM TYPE: Floppy disk		STREET: 6300 Columbia Center, 701 Fifth Avenue	
COMPUTER: IBM PC compatible		CITY: Seattle	
OPERATING SYSTEM: PC-DOS/MS-DOS		STATE: Washington	
CURRENT APPLICATION DATA:		COUNTRY: USA	
APPLICATION NUMBER: US/09/072,596		ZIP: 98104-7092	
FILING DATE: 05-MAY-1998		CLASSIFICATION:	
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND		ATTORNEY/AGENT INFORMATION:	
NUMBER OF SEQUENCES: 241		NAME: Maki, David J.	
CORRESPONDENCE ADDRESS:		REGISTRATION NUMBER: 31,392	
ADDRESSEE: SPEED and BERRY LLP		REFERENCE/DOCKET NUMBER: 210121.417C6	
STREET: 6300 Columbia Center, 701 Fifth Avenue		TELECOMMUNICATION INFORMATION:	
CITY: Seattle			
TREATMENT:			

TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 99:  
SEQUENCE CHARACTERISTICS:

LENGTH: 51 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-09-072-596-99

RESULT 5  
US-08-465-640-2  
Sequence 2, Application US/08465640  
Patent No. 5935077  
GENERAL INFORMATION:  
APPLICANT: ANDERSEN, Peter  
APPLICANT: ANDERSEN, Ase Bengaard  
APPLICANT: HASLEV, Kaare  
APPLICANT: SØRENSEN, Anne Lund  
TITLE OF INVENTION: TUBERCULOSIS VACCINE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,640  
FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,182  
FILING DATE: 20-SEP-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCW/DK94/00273  
FILING DATE: 01-JUL-1994

ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28, 005  
REFERENCE/DOCKET NUMBER: ANDERSEN=3A

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 24633

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 95 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein

US-08-465-640-2

Db 6 WNFAGTERRASAATQG 20

RESULT 6  
US-09-368-590-2  
Sequence 2, Application US/09368590  
Patent No. 6187563  
GENERAL INFORMATION:  
APPLICANT: Solimena, Michel  
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR  
FILE REFERENCE: 101918-200 (OCR-941)  
CURRENT APPLICATION NUMBER: US/09/368, 590  
CURRENT FILING DATE: 1999-08-04  
EARLIER APPLICATION NUMBER: 60/095, 657  
EARLIER FILING DATE: 1998-08-07  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FASTSEQ for Windows version 3.0  
SEQ ID NO 2: 2293  
LENGTH: 2293  
TYPE: PRT  
ORGANISM: Human

RESULT 7  
US-09-202-893B-4  
Sequence 4, Application US/09202893B  
Patent No. 6319692  
GENERAL INFORMATION:  
APPLICANT: KADOTA, Mariko  
APPLICANT: KIWAKI, Mayumi  
APPLICANT: SAWAKI, Saeko  
APPLICANT: SHIRASAWA, Yukiko  
APPLICANT: SONE, Harue  
APPLICANT: SAKO, Tomoyuki

TITLE OF INVENTION: METHODS FOR TRANSFERRING GENE INTO CHROMOSOME  
FILE REFERENCE: 980794/PG  
CURRENT APPLICATION NUMBER: US/09/202, 893B

CURRENT FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: PCP/JP97/02187  
PRIOR FILING DATE: 1997-06-25  
PRIOR APPLICATION NUMBER: JP 8/184266  
PRIOR FILING DATE: 1996-06-26  
PRIOR APPLICATION NUMBER: JP 8/257764  
PRIOR FILING DATE: 1996-09-06  
NUMBER OF SEQ ID NOS: 9  
SEQ ID NO 4  
LENGTH: 418  
TYPE: PRT  
ORGANISM: Lactobacillus casei

US-09-202-893B-4

Query Match 49.4%; Score 38; DB 4; Length 418;  
Best Local Similarity 61.5%; Pred. No. 62; Mismatches 3; Indels 0; Gaps 0;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

RESULT 8  
US-09-509-902A-9  
Sequence 9, Application US/09509902A

Query Match 100.0%; Score 77; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 11  
 US-08-478-572-11  
 Sequence 11, Application US/08478572  
 Patent No. 5568526  
 GENERAL INFORMATION:  
     APPLICANT: Garman, Richard  
     APPLICANT: Greenstein, Julia  
     APPLICANT: Kuo, Mei-chang  
     APPLICANT: Rogers, Bruce  
     APPLICANT: Franzen, Henry  
     APPLICANT: Chen, Xian  
     APPLICANT: Evans, Sean  
     APPLICANT: Shaked, Ze'ev  
     TITLE OF INVENTION: T CELL EPITOPESES OF THE MAJOR ALLERGENS FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
     TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
     NUMBER OF SEQUENCES: 207  
     CORRESPONDENCE ADDRESS:  
         ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
         STREET: 610 LINCOLN STREET  
         CITY: WALTHAM  
         STATE: MA  
         COUNTRY: USA  
         ZIP: 01154  
     COMPUTER READABLE FORM:  
         MEDIUM TYPE: Floppy disk  
         COMPUTER: IBM PC compatible  
         OPERATING SYSTEM: PC-DOS/MS-DOS  
         SOFTWARE: ASCII TEXT  
 CURRENT APPLICATION DATA:  
     APPLICATION NUMBER: US/08/478,572  
     FILING DATE: 07-June-1995  
     CLASSIFICATION:  
         PRIORITY APPLICATION DATA:  
             APPLICATION NUMBER: 08/445,307  
             ATTORNEY/AGENT INFORMATION:  
                 NAME: CRAIG, ANNE I.  
                 REGISTRATION NUMBER: 32,976  
                 REFERENCE/DOCKET NUMBER: 017.6US  
             TELECOMMUNICATION INFORMATION:  
                 TELEPHONE: (617) 466-6000  
                 TELEFAX: (617) 466-6040  
             INFORMATION FOR SEQ ID NO: 11:  
                 SEQUENCE CHARACTERISTICS:  
                     LENGTH: 29 amino acids  
                     TYPE: amino acid  
                     TOPOLOGY: linear  
                     MOLECULE TYPE: peptide  
                     FRAGMENT TYPE: N-terminal  
             US-08-478-572-11  
     Query Match           Score 35; DB 2; Length 29;  
     Best Local Similarity 45.5%; Pred. No. 9.2;  
     Matches            50.0%; 2; Mismatches 4; Indels 0; Gaps 0;  
     Qy                1 WNFAGTIAASA 12  
     Db                15 WAFSGVAAATESA 26  
  
 RESULT 12  
 US-08-478-572-74  
 Sequence 74, Application US/08478572  
 Patent No. 5568526  
 GENERAL INFORMATION:  
     APPLICANT: Garman, Richard  
     APPLICANT: Greenstein, Julia  
     APPLICANT: Kuo, Mei-chang  
     APPLICANT: Rogers, Bruce  
     APPLICANT: Franzen, Henry  
     APPLICANT: Chen, Xian  
     APPLICANT: Evans, Sean  
     APPLICANT: Shaked, Ze'ev  
     TITLE OF INVENTION: T CELL EPITOPESES OF THE MAJOR ALLERGENS FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
     TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
     NUMBER OF SEQUENCES: 207  
     CORRESPONDENCE ADDRESS:  
         ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
         STREET: 610 LINCOLN STREET  
         CITY: WALTHAM  
         STATE: MA  
         COUNTRY: USA  
         ZIP: 01154  
     COMPUTER READABLE FORM:  
         MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII TEXT  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/484, 296  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/445, 307  
 FILING DATE: 07 June 1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CRAIG, ANNE I.  
 REGISTRATION NUMBER: 32, 976  
 REFERENCE/DOCKET NUMBER: 017.6US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 466-6000  
 TELEX/FAX: (617) 466-6040  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: N-terminal  
 US-08-484-296-11

Query Match 45.5%; Score 35; DB 4; Length 29;  
 Best Local Similarity 50.0%; Pred. No. 9.2; Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 WNFAGIERRASA 12  
 | : | : | |  
 Db 15 WAFSGVAVATESA 26

RESULT 14  
 Sequence 74, Application US/08484296  
 Patent No. 626841  
 GENERAL INFORMATION:  
 APPLICANT: Garman, Richard  
 APPLICANT: Greenstein, Julia  
 APPLICANT: Kuo, Mei-chang  
 APPLICANT: Rogers, Bruce  
 APPLICANT: Franzen, Henry  
 APPLICANT: Chen, Xian  
 APPLICANT: Evans, Sean  
 APPLICANT: Shaked, Ze'ev  
 TITLE OF INVENTION: T CELL EPITOPEs OF THE MAJOR ALLERGENs  
 NUMBER OF SEQUENCES: 207  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION  
 STREET: 610 LINCOLN STREET  
 CITY: WALTHAM  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII TEXT  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/484, 296  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/445, 307  
 FILING DATE: 07 June 1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CRAIG, ANNE I.  
 REGISTRATION NUMBER: 32, 976

RESULT 15  
 PCT-US95-04481-3  
 Sequence 3, Application PC/US9504481  
 GENERAL INFORMATION:  
 APPLICANT: TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Dust M  
 NUMBER OF SEQUENCES: 54  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/04481  
 FILING DATE:  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/227, 772  
 FILING DATE: April 14, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vanstone, Darlene A.  
 REGISTRATION NUMBER: 35, 279  
 REFERENCE/DOCKET NUMBER: 017.5 PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 466-6000  
 TELEX/FAX: (617) 466-6040  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 29 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 PCT-US95-04481-3

Query Match 45.5%; Score 35; DB 5; Length 29;  
 Best Local Similarity 50.0%; Pred. No. 9.2; Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 WNFAGIERRASA 12  
 | : | : | |  
 Db 15 WAFSGVAVATESA 26

Search completed: July 3, 2003, 14:07:46  
 Job time: 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 3, 2003, 14:03:09 ; Search time 24 Seconds

(without alignments)  
128.779 Million cell updates/sec

Title:

US-09-830-839-6

Perfect score:

77

Sequence: 1 WNFAGIEAASAIQ 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters:

671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21;\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rabbit:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_ivirus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	47	61.0	714	16 Q9CES5
2	47	225	2 Q9RF6	Q9RF6 rhodobacter
3	42.5	55.2	141	16 Q9HYR3
4	42	369	16 Q9ZB39	Q9ZB39 chlamydia p
5	42	2030	10 Q8RZC9	Q8RZC9 oryza sativ
6	41	53.2	440	2 Q9FAD1
7	41	53.2	443	2 Q9AMK7
8	41	53.2	445	16 Q9KV76
9	41	53.2	540	16 Q9CGX6
10	41	53.2	803	5 Q9N978
11	40	51.9	134	8 Q9XMD0
12	40	51.9	134	8 Q9XMD1
13	40	51.9	323	16 Q8UH31
14	40	51.9	330	17 Q27223
15	40	51.9	379	8 Q9B1NO
16	40	51.9	379	8 Q9B106

RESULT 1	Q9CESS	PRELIMINARY	PRT	714 AA.
ID	Q9CES5			
AC	Q9CES5;			
DT	01-JUN-2001 ("TREMBLrel. 17, Last sequence update")			
DT	01-JUN-2002 ("TREMBLrel. 21, Last annotation update")			
DE	Glutamine transporter permease and substrate binding protein			
DE	protein.			
GN	GLNP OR LLI1759.			
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).			
OC	Bacteria; Firmicutes; Lactobacillales;			
OC	Streptococaceae; Lactococcaceae.			
OX	NCBI_TAXID=1360;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LIL403;			
RX	MEDLINE="21235186; PubMed=11337471;"			
RA	Boletin A., Winkler P., Mauger S., Jaillon O., Malarme K., Weissenbach J., Ehrlich S.D., Sorkin A.;			
RA	"The complete genome sequence of the lactic acid bacterium Lactococcus			
RT	lactis spp. lactis LIL403.";			
RL	Genome Res. 11:731-753 (2001).			
CC	-1- FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY). SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEMS.			
CC	-1- PROTEIN-DIRECTED TRANSPORT SYSTEMS.			
DR	EMBL: AE06405; AAC05857.1; -			
DR	HSPP; P10344; IWDN; InterPro: IPR00515; BPD_transp.			
DR	InterPro: IPR001311; SBP_glu_receptor.			
DR	InterPro: IPR01638; SBP_bac_3.			
DR	Pfam: PF00528; BPD_transp; 1.			
DR	Pfam: PF00497; SBP_bac_3; 2.			
DR	SMART: SM00627; PBP_2.			
DR	PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.			
DR	PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.			

KW	transmembrane; Transport; Complete proteome.
SQ	SEQUENCE 714 AA; 78377 MW; 4EBEBE5545A968EB CRC64;
Query Match	61.0%; Score 47; DB 16; Length 714;
Best Local Similarity	50.0%; Pred. No. 19;
Matches	7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Qy	1 WNFAGIEAAASAIQ 14
Db	303 WNFIGFQAAVDSVQ 315
RESULT 2	
O9REF6	PRELIMINARY; PRT; 225 AA.
ID	O9REF6
AC	O9REF6;
DT	01-MAY-2000 (TREMBrel. 13, Created)
DT	01-MAY-2000 (TREMBrel. 13, Last sequence update)
DT	01-MAR-2002 (TREMBrel. 20, Last annotation update)
DR	URED.
OS	Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC	Rhodobacter.
OX	NCBI_TaxID:1063;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=2.4.1;
RA	Choudhary M., Kaplan S.;
RT	"DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.4.1.";
RL	Nucleic Acids Res. 0; 0-0 (2000).
DR	EMBL: AF195122; AAF24251.1; -.
DR	InterPro; IPR002669; URED.
DR	Pham; PF01774; URED; 1.
SO	SEQUENCE 225 AA; 23417 MW; 59EC44A05802C29F CRC64;
Query Match	57.1%; Score 44; DB 2; Length 225;
Best Local Similarity	75.0%; Pred. No. 18;
Matches	9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy	4 AGIEAAASAIQ 15
Db	178 AGVEAAASAFDG 189
RESULT 3	
O9HYR3	PRELIMINARY; PRT; 141 AA.
ID	O9HYR3
AC	O9HYR3;
DT	01-MAR-2001 (TREMBrel. 16, Created)
DT	01-MAR-2001 (TREMBrel. 16, Last sequence update)
DT	01-OCT-2001 (TREMBrel. 18, Last annotation update)
DE	Hypothetical protein PA3332.
GN	PA3332.
OS	Pseudomonas aeruginosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
OX	NCBI_TaxID:287;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=J138;
RX	MEDLINE=20330349; PubMed=10871362;
RA	Shirai M., Hikakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.;
RA	"Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA.";
RT	Nucleic Acids Res. 28:2311-2314(2000).
RL	EMBL; AB001636; AAD18653.1; -.
DR	EMBL; AE002184; AAF38105.1; -.
DR	EMBL; AP002546; BAA98719.1; -.
DR	TIGR; CP0240; -.
DR	InterPro; IPR005244; Cons_hypoth423.
DR	TIGRFAMs; TIGR00423; TIGR00423; 1.
KW	Complete proteome.
SO	SEQUENCE 369 AA; 41556 MW; 05AC33C15479B192 CRC64;
Query Match	54.5%; Score 42; DB 16; Length 369;
Best Local Similarity	77.8%; Pred. No. 65;
Matches	7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy	1 WNFAGIEA 9
Db	305 WNFAGIEA 313
RESULT 5	
OBRZC9	PRELIMINARY; PRT; 2030 AA.
ID	OBRZC9

"Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.";  
Nature 406:959-964 (2000).  
EMBL: AE004755; AA0G05720.1; -.  
DR Hypothetical protein; Complete proteome.

AC 08RZC9; DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE P0510C12.24 protein.  
 RN [1]—  
 OS *Oryza sativa* (*Japonica* cultivar-group).  
 OC *Bukaryota*; *Viridiplantae*; *Streptophyta*; *Embryophyta*; *Tracheophyta*;  
 OC *Spermato phyta*; *Magnoliophyta*; *Liliopsida*; *Poales*; *Poaceae*;  
 OC *Ehrhartioideae*; *Oryzeae*; *Oryza*.  
 DR NCBI\_TaxID=39947;  
 RN [1]—  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV; NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3)" genomic DNA, chromosome 1, PAC  
 RT clone:P0510C12;"  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF003725; BAB50479\_1; MW; 26C0424D2D4867ED CRC64;  
 SEQUENCE 2030 AA; 223434 MW;  
 Query Match 54.5%; Score 42; DB 10; Length 2030;  
 Best Local Similarity 60.0%; Pred. No. 3.8e+02;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASIQG 15  
 || :| | :|| |  
 Db 400 WNSSGSQAASRAPG 414

RESULT 6

ID 09FAD1 PRELIMINARY; PRT; 440 AA.  
 AC 09FAD1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE CADB.  
 GN Moritella japonica.  
 OS Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
 OC Moritella.  
 OX NCBL\_TAXID=89067;  
 RN [1]—  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE-20406833; PubMed-10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ernolatova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
 cholerae.", Nature 406:477-483 (2000).  
 RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB041735; BAB16717\_1;  
 DR InterPro; IPR002293; AA\_REL\_PRIMEASE1.  
 DR InterPro; IPR004841; Permase.  
 DR Pfam; PF00324; aa\_permease; 1.  
 SQ SEQUENCE 440 AA; 45577 MW; DBDE4CADR8958CD9 CRC64;

Query Match 53.2%; Score 41; DB 2; Length 440;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAA 10  
 || :| | :|| |  
 Db 199 WSFVGYESAA 208

RESULT 7

ID 09AMK7 PRELIMINARY; PRT; 443 AA.  
 AC 09AMK7;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 20, Last annotation update)  
 DE Cadaverine/Lysine antiporter.

AC 08RZC9; DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; gamma subdivision; vibrionaceae; vibrio.  
 OC NCBI\_TaxID=672;  
 RN [1]—  
 RP SEQUENCE FROM N.A.  
 RA Rhee J., Lee J., Lee S., Chung S., Rhee J., Choi S., Ryu P.;  
 RT "Vibrio vulnificus cadBA genes.", Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.  
 RL EMBL; AF324470; AAK01663\_1; -.  
 DR Intertro; IPR002293; AA\_REL\_PRIMEASE1.  
 DR Intertro; IPR004841; Permase.  
 DR Pfam; PF00324; aa\_permease; 1.  
 SQ SEQUENCE 443 AA; 46752 MW; B8FB7E188036751 CRC64;

Query Match 53.2%; Score 41; DB 2; Length 443;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAA 10  
 || :| | :|| |  
 Db 199 WSFVGYESAA 208

RESULT 8

ID 09KV76 PRELIMINARY; PRT; 445 AA.  
 AC 09KV76;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 21, Last sequence update)  
 DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cadaverine/Lysine antiporter CadB, putative.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; vibrionaceae; vibrio.  
 OX NCBI\_TAXID=666;  
 RN [1]—  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE-20406833; PubMed-10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ernolatova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
 cholerae.", Nature 406:477-483 (2000).  
 RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB041116; AAF93455\_1; -.  
 DR TIGR; VC0280;  
 DR Intertro; IPR002293; AA\_REL\_PRIMEASE1.  
 DR Intertro; IPR004841; Permase.  
 DR Pfam; PF00324; aa\_permease; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 445 AA; 46907 MW; 3A64F21B9680684C CRC64;

Query Match 53.2%; Score 41; DB 16; Length 445;  
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAA 10  
 || :| | :|| |  
 Db 199 WSFVGYESAA 208

RESULT 9

ID 09CGX6 PRELIMINARY; PRT; 540 AA.  
 AC 09CGX6;  
 ID 09CGX6;  
 AC 09CGX6;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2002 (TREMbrel. 20, Last annotation update)

DE ABC transporter permease protein.

GN YJJD OR LL0956.

OS Lactococcus lactis (ssp. *lactis*) (*Streptococcus lactis*)

OC Bacteria; Firmicutes; Bacilli/Clostridium group; Lactobacillales;

OC Streptococcaceae; Lactococcaceae.

OX NCBI\_TAXID=1360;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TU1403;

RX MEDLINE=21235186; PubMed=11337471;

RA Bolotin A.; Wincker P.; Mauzer S.; Jallion O.; Malarrie K.,

RA Weissenbach J.; Birlich S.D.; Sorokin A.;

RT "The complete genome sequence of the lactic acid bacterium *Lactococcus* *lactis* ssp. *lactis* TU1403.";

RL Genome Res. 11:731-753(2001).

DR EMBL; AR00531; ANK05064; 1; -

KW Complete proteome.

SO SEQUENCE 540 AA; 60283 MW; E896CF783987560E CRC64;

Query Match 53.2%; Score 41; DB 16; Length 540;

Best Local Similarity 69.2%; Pred. No. 1.4e+02;

Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0; CC 0;

Db 3 FAGIEAAASAIQ 15  
339 FVGNETAASAIRG 351

RESULT 10

Q9N978 PRELIMINARY; PRT; 803 AA.

AC Q9N978

DT 01-OCT-2000 (TREMbrel. 15, Created)

DT 01-OCT-2000 (TREMbrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMbrel. 21, Last annotation update)

DE Possible amino acid transporter.

GN PI295.10.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI\_TAXID=5664;

RN [1]

RA Mesuy D.; Purnelle B.; Goffeau A.; Ivens A.C.; Quail M.,

RA Rajandream M.A.; Barrell B.G.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=PRIEDLIN;

RX MEDLINE=9816435; PubMed=9477341;

RA Ivens A.C.; Lewis S.M.; Bagherzadeh A.; Zhang L.; Chan H.M.,

RA Smith D.F.;

RT "A physical map of the *Leishmania major* Friedlin genome.";

RL Genome Res. 8:135-145(1998);

DR EMBL; AU359773; CAB95218; 1; -

DR InterPro; IPR002933; AA\_REL\_PRMEASEL.

DR InterPro; IPR004811; Permease.

DR Pfam; PF00324; aa\_permeases; 1.

SO SEQUENCE 803 AA; 84779 MW; 0AEBD76516440558 CRC64;

Query Match 53.2%; Score 41; DB 5; Length 803;

Best Local Similarity 42.9%; Pred. No 2.1e+02;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0; CC 0;

Db 285 WNYANIEANGAVVE 298

RESULT 11

Q9XMD0 PRELIMINARY; PRT; 134 AA.

AC Q9XMD0

DT 01-NOV-1999 (TREMbrel. 12, Created)

DT 01-NOV-1999 (TREMbrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMbrel. 19, Last annotation update)

DE Cytochrome b (Fragment).

OS Pteronotus parnellii (Parnell's mustached bat).

OG Mitochondrion.

OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae; Pteronotus.

OC Pteronotus.

OX NCBI\_TAXID=59476;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20072938; PubMed=10603267;

RA Kennedy M.; Patterson A.M.; Morales J.C.; Parsons S.; Winnington A.P.,

RA Spencer H.G.;

RA "The long and short of it: Branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*.";

RL Mol. Phylogenet. Evol. 13:405-416(1999).

CC -1 - FUNCTION: COMPLEX OF THE UBIQUINOL-CYTOCHROME C REDUCTASE CYTOCHROME b (Fragment).

CC "The long and short of it: Branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*.";

CC Mol. Phylogenet. Evol. 13:405-416(1999).

CC -1 - FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS (BY SIMILARITY).

CC -1 - COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN (BY SIMILARITY).

CC -1 - SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).

CC -1 - SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

DR EMBL; AR14407; AAD34991; -

DR InterPro; IPR001719; Cyt\_b\_N; 1;

DR Pfam; PF00033; cytochrome\_c\_b\_N; 1;

DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.

KW Electron transport; Heme; Mitochondrion; Respiratory chain; Transmembrane; PT NOLTER 134 134 AA; 15054 MW; 54FA6146B46B3C03 CRC64;

Query Match 51.9%; Score 40; DB 8; Length 134;

Best Local Similarity 50.0%; Pred. No. 50;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0; CC 0;

Db 31 WNFQSSLANCLAVQ 44

RESULT 12

Q9XMD1 PRELIMINARY; PRT; 134 AA.

AC Q9XMD1

DT 01-NOV-1999 (TREMbrel. 12, Created)

DT 01-NOV-1999 (TREMbrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMbrel. 19, Last annotation update)

DE Cytochrome b (Fragment).

OS Pteronotus parnellii (Parnell's mustached bat).

OG Mitochondrion.

OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae; Pteronotus.

OC Pteronotus.

OX NCBI\_TAXID=59476;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20072938; PubMed=10603267;

RA Kennedy M.; Patterson A.M.; Morales J.C.; Parsons S.; Winnington A.P.,

RA Spencer H.G.;

RA "The long and short of it: Branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*.";

RL Mol. Phylogenet. Evol. 13:405-416(1999).

CC -1 - FUNCTION: COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY). WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL; AF44071; AAD55000\_1; -.  
 DR InterPro; IPR000179; Cyt\_b\_b6;  
 DR Pfam; PF00033; Cytochrome\_b\_N; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 KW Electron transport; Heme; Mitochondrion; Respiratory chain;  
 KW Transmembrane;  
 FT NON\_TER 134 134  
 SQ SEQUENCE 134 AA; 15086 MW; 5068D4E9A6D23C4 CRC64;  
 Query Match Similarity 51.9%; Score 40; DB 8; Length 134;  
 Best Local Similarity 50.0%; Pred. No. 50;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGTEAASAQI 14  
 RP 1||| : ||| : ||| 44  
 Db 31 WNFGSLLACLAQV 44

RESULT 13

PRELIMINARY; PRT; 323 AA.

ID Q8UH81 PRELIMINARY; PRT; 323 AA.

AC Q8UH81; 2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)

DE Hypothetical protein Atuo0802.

GN ATU0802 OR AGR.C\_168.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970);  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TAXID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2160850; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Woo L.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovée D. Srl., Grant C.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Bush D.,  
 RA Kuttrabin T., Levy M.-J., Maclelland E., Palmieri A.,  
 RA Raymond C., Reuse G., Saenphimmaachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2160851; PubMed=11743194;  
 RA Goodier B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,  
 RA Ourolo B., Goldman B.S., Cao Y., Askarazi M., Halling C., Mullin L.,  
 RA Hounsell K., Gordon J., Vaudin M., Tartchouk O., Epp A., Liu F.,  
 RA William C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the Plant Pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 RL [3]  
 DR EMBL; AE000813; AAB86611\_1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 323 AA; 35066 MW; C2097F09R00C744E CRC64;

Query Match Similarity 51.9%; Score 40; DB 16; Length 323;  
 Best Local Similarity 61.5%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 5; 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 FAGIEAAASAQG 15  
 RP |||| : ||| : ||| 15  
 Db 11 FAGISCATALOG 23

RESULT 14

PRELIMINARY; PRT; 330 AA.

ID Q27223 PRELIMINARY; PRT; 330 AA.

AC Q27223; 1998 (TREMBLrel. 05, created)  
 DT 01-JUN-1998 (TREMBLrel. 05, last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE NA+/CA+ EXCHANGING protein related.

GN MTH1155.

OS Methanobacterium thermoaotrophicum.

OC Archaea; Buryarchaeota; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 OX NCBI\_TAXID=18420;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DELTA\_H;  
 RX MEDLINE=98037514; Pubmed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., Derooughery C., Lee H.-M., Dubois J.,  
 RA Aldridge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiwan N., Caruso A., Bush D., Safer H., Pawlowski D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,  
 RA "Complete genome sequence of Methanobacterium thermoautrophicum  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL; AE000815; AAB05644\_1; -.  
 DR InterPro; IPR004481; K\_NaCaExchang.  
 DR InterPro; IPR004837; NaCa\_Exchang.  
 DR Pfam; PF01699; Na\_Ca\_Ex; 2.  
 DR TIGRFAMS; TIGR0367; K\_NaCaexchang-rel; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 330 AA; 34762 MW; BEETB0295FA5A4DF CRC64;

Query Match Similarity 46.7%; Score 40; DB 17; Length 330;  
 Best Local Similarity 46.7%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WNFAGTEAASAQI 15  
 RP 1||| : ||| : ||| 15  
 Db 95 WNIAGLIGLGISSWVG 109

RESULT 15

OBIN0 PRELIMINARY; PRT; 379 AA.

ID OBIN0 PRELIMINARY; PRT; 379 AA.

AC OBIN0; 2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 18, Last annotation update)

DE Cytochrome b.  
 GN CYTB.  
 OS Pteronotus davyi (Davy's naked-backed bat).  
 OG Mitochondrion.  
 OC Bukarvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Mormoopidae;  
 OC Pteronotus.  
 OX NCBITAXID=94956;

[1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=21542; AND 35483;  
 RA Lewis-Orritt N., Porter C.A., Baker R.J.;  
 RT "Molecular Systematics of the Family Mormoopidae (Chiroptera) Based on  
 RT Mitochondrial cytochrome b and Nuclear RAG-2 Gene Sequences.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases;  
 CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

CC: COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC: RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC: COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC: -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC: BOUND TO THE PROTEIN (BY SIMILARITY).  
CC: -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC: CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
CC: -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
DR: EMBL; AP338672; AAK21932.1; -  
DR: EMBL; AP338670; AAK21930.1; -  
DR: InterPro; IPR00179; Cyt\_b\_b6;  
DR: Pfam; PF00032; cytochrome\_b\_C; 1.  
DR: Pfam; PF00033; cytochrome\_b\_N; 1.  
DR: PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR: PROSITE; PS00193; CYTOCHROME\_B\_QO; UNKNOWN\_1.  
KW: Electron transport; Heme; Mitochondrion; Respiratory chain;  
KW: Transmembrane.  
SEQUENCE 379 AA: 42664 MW: 27FC6661E666E624 CRG64;

Query Match 51.9% Score 40; DB 8; Length 379;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 WNFAGIEAAASAIQ 14  
||| : || | : |  
Db 31 WNFAGIALLAACLAVQ 44

Search completed: July 3, 2003, 14:06:48

Job time : 25 secs

Copyright (c) 1993 - 2003 Compugen Ltd.

### OM protein - protein search, using sw model

Run on: July 3, 2003, 14:02:14 ; Search time 6.5 Seconds  
 (without alignments)  
 95.715 Million cell updates/sec

Title: US-09-830-839-6

Perfect score: 77  
 Sequence: 1 WNFAGIEAAASAIQG 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	77	100.0	94	1 ESA6_MYCTU
2	43	55.8	476	1 NOEA_RHIME
3	41	53.2	384	1 Fhub_BACSU
4	40	51.9	294	1 G3F_ESCBL
5	40	51.9	2562	1 SPCQ_HUMAN
6	39	50.6	370	1 CYB_EPICE
7	39	50.6	379	1 CYB_ZALICA
8	39	50.6	425	1 ERWCH
9	39	50.6	445	1 YJDE_ECOLI
10	39	50.6	990	1 TNP7_ECOLI
11	38	49.4	95	1 ESA6_NYCLC
12	38	49.4	444	1 CADB_ECOLI
13	38	49.4	923	1 YAU_SCHDO
14	38	49.4	962	1 ARVC_HUMAN
15	38	49.4	969	1 ARVC_MOUSE
16	38	49.4	988	1 TNP2_ECOLI
17	37	48.1	75	1 OCTI_MACEU
18	37	48.1	196	1 RBMS_XENPO
19	37	48.1	335	1 YD59_SCIRPO
20	37	48.1	379	1 CYB_CHEMA
21	37	48.1	404	1 CYB_ZAPTR
22	37	48.1	475	1 DG1A_RHITO
23	37	48.1	492	1 HUPR_RHOCIA
24	37	48.1	507	1 PSBB_SYN3
25	37	48.1	509	1 PSBB_ANASP
26	37	48.1	509	1 PSBB_CYAPA
27	37	48.1	660	1 PBBB_CYAPA
28	37	48.1	660	1 PAB3_ARATH
29	36	46.8	125	1 CICP_RHOPA
30	36	46.8	145	1 Y008_MYCTU
31	36	46.8	211	1 MT04_SYN3
32	36	46.8	376	1 FDH_ORYSA
33	36	46.8	377	1 FDH_HORVU

### ALIGNMENTS

RESULT 1	ES64_MYCTU	ID	ES64_MYCTU	STANDARD;	PRT:	94 AA.
		AC	057165; 084901;			
		DT	30-MAY-2000 (Rel. 39, Created)			
		DT	30-MAY-2000 (Rel. 39, last sequence update)			
		DT	15-JUN-2002 (Rel. 41, last annotation update)			
		DE	6 kDa early secretory antigenic target (ESAT-6).			
		GN	ESAT6 OR RV3815 OR MT3999 OR Mtv027.10.			
		OS	Mycobacterium tuberculosis, and			
		OS	Mycobacterium bovis.			
		OC	Bacteria; Actinobacteria; Actinomycetaceae; Corynebacteriaceae; Mycobacteriidae; Actinomycetales.			
		OX	NCBI_TaxID=1773, 1765;			
		RN	[1]			
		RP	SEQUENCE FROM N.A.			
		RC	SPECIES=M_tuberculosis; STRAIN=H37Rv;			
		RX	MEDLINE=95247251; PubMed=7729876;			
		RA	Sorensen A.L., Nagai S., Holten G., Anderson P., Anderson A.B.,			
		RA	Gordon S.V., Biglinier K., Gas S., Barry C.E. III, Tekla F.,			
		RT	Chillingworth T., Connor R., Badcock K., Basham D., Brown D., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutister S., Seeger K., Shelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
		RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutister S., Seeger K., Shelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
		RT	"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";			
		RT	Nature 393:537-544 (1998).			
		RL				
		RN				
		RP	SEQUENCE FROM N.A.			
		RC	SPECIES=M_tuberculosis; STRAIN=CDC 1551 / Oshkosh;			
		RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayan L.A., Ernolalaya M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bisbal W.,			
		RA	"Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";			
		RL	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.			





Matches	7; Conservative	1; Mismatches	4; Indels	0; Gaps	0;
Qy	1 WNFAGIEAAASA 12				
Db	70 WNEAGVEVVAE 81				
RESULT 5					
SPCC_HUMAN	STANDARD; PRT; 2564 AA.				
ID SPCC_HUMAN	PRT; 2564 AA.				
AC Q9H254; Q9HC0D; Q9H3G8; Q9H1K7; Q9H1K8; Q9H1K9;					
DT 15-JUN-2002 (Rel. 41; Last sequence update)					
DT 15-JUN-2002 (Rel. 41; Last annotation update)					
DE Spectrin beta chain, brain 3 (Spectrin, non-erythrocytic spectrin, beta-IV spectrin).					
GN SP1BN4 OR SPB1N3 OR KIAA1642.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. OX NEBI_TAXID=9606;					
RN [1]					
SEQUENCE FROM N.A. (ISOFORM 1); MEDLINE=21316449; Published=11294630;					
RA Tse W.T., Tang J., Jin O., Korngren C., John K.M., Kung A.L., Gwynn B., Peters L.L., Lux S.E.;					
RT "A new spectrin, beta-IV, has a major truncated isoform that associates with promyelocytic leukemia protein nuclear bodies and the nuclear matrix.";					
RT J. Biol. Chem. 276:23974-23985(2001). [2]					
RP SEQUENCE FROM N.A. (ISOFORMS 1 TO 4); MEDLINE=20539976; Published=11086001;					
RA Berghs S., Aguzzi D., Dirkx R., Maksimova E., Stabach P., Hermel J.-M., Zhang J.-P., Philbrick W., Slepnev V., Ort T., Solimeni M.;					
RT "BetaIV spectrin, a new spectrin localized at axon initial segments and nodes of ranvier in the central and peripheral nervous system.";					
RL J. Cell Biol. 151:985-1002(2000). [3]					
RN SEQUENCE OF 386-2382 FROM N.A. (ISOFORM 1).					
RP TISSUE-BRAIN; RX MEDLINE=20436683; Published=1097877;					
RA Nagase T., Kikuno R., Nakayama M., Hirokawa N., Ohara O.; RT "Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";					
RL DNA Res. 7:273-291(2000). [4]					
CC ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; are produced by alternative splicing.					
CC - TISSUE SPECIFICITY: Abundantly expressed in brain and pancreatic islets.					
CC - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.					
CC - SIMILARITY: CONTAINS 2 CALPONTIN-HOMOLOGY (CH) DOMAINS.					
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.					
CC - SIMILARITY: CONTAINS 18 SPECTRIN REPEATS.					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).					
DR EMBL; AF0311855; AAG44473.1; -					
DR EMBL; AF082075; AGG38874.1; -					
DR EMBL; AY004226; AAF93172.1; -					
DR EMBL; AY04227; AAF93173.1; -					
DR EMBL; AB048862; BAB13468.1; -					
DR HSSP; Q01082; 1BKR.					
DR Genew; HGNC:14890; SP1BN4.					
RESULT 6					
CYB_EPICE	Best Local Similarity: 42.9%; Pred. No. 1.4e+02; Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;				
ID CYB_EPICE	STANDARD; PRT; 370 AA.				
AC Q48043; DT 16-OCT-2001 (Rel. 40; Created)					
DT 16-Oct-2001 (Rel. 40; Last sequence update)					
DR Pfam; PF00169; PH; 3; DR Pfam; PF00307; CH; 2; DR InterPro; IPR01715; Calponin-like. DR InterPro; IPR01849; PH. DR InterPro; IPR02017; Spectrin. DR InterPro; IPR01605; Spectrin.PH. DR Pfam; PF00435; spectrin; 56. DR PRINTS; PRO0683; SPECTRINPH. SMART; SM0033; CH; 2. DR PROSITE; PS00019; ACTIN_N_1; 1. DR PROSITE; PS00020; ACTIN_N_2; 1. DR PROSITE; PS00021; CH; 2. DR PROSITE; SS0003; PH_DOMAIN; 1. DR CYTOSKELETON; Membrane; Repeat; Actin-binding; Capping protein; KW Alternative splicing. FT DOMAIN 1 282 FT DOMAIN 61 165 FT DOMAIN 180 282 FT REPEAT 309 354 FT REPEAT 398 419 FT REPEAT 429 533 FT REPEAT 535 642 FT REPEAT 644 771 FT REPEAT 773 879 FT REPEAT 881 985 FT REPEAT 1019 1086 FT REPEAT 1088 1197 FT REPEAT 1199 1303 FT REPEAT 1305 1408 FT REPEAT 1410 1513 FT REPEAT 1515 1619 FT REPEAT 1621 1725 FT REPEAT 1727 1832 FT REPEAT 1834 1940 FT REPEAT 1942 2046 FT REPEAT 2048 2107 FT DOMAIN 2418 2527 FT VARSPlic 2151 2157 FT VARSPlic 1258 1286 FT VARSPlic 1287 1309 FT VARSPlic 1310 2564 FT VARSPlic 2113 2154 FT VARSPlic 2155 2564 FT VARSPlic 604 608 FT CONFLICT 714 714 FT CONFLICT 1189 1189 FT CONFLICT 1193 1193 FT CONFLICT 1331 1331 FT CONFLICT 2564 AA; 288982 MW; 5202E7D11D601ECC CRC64; SQ SEQUENCE					
Query Match	51.9%; Score 40; DB 1; Length 2564;				
Best Local Similarity	42.9%; Pred. No. 1.4e+02;				
Matches	6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;				
Qy	1 WNFAGIEAAASAQ 14				
Db	1022 WRSLSGEALQALE 035				

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cytochrome B.

GN MTCYB OR COB OR CYTB.

OS Epicrates cenchria.

OC Mitochondrion.

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Henophidia; Boidae;

OC Boinae; Epicrates.

OC NCBI\_TaxID=51743;

RN [1]

RE SEQUENCE FROM N.A.

RA Campbell B.N.;

PRT "HIC Sunt Serpentes - molecular phylogenetics and the Boidae  
(Serpentes: Boidae).";

RU Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.

CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS (BY SIMILARITY).

CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN (BY SIMILARITY).

CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR HSSP; P00157; 1BB3.

CC DR InterPro; IPR000179; Cyt\_b\_b6.

CC DR Pfam; PF00032; cytochrome\_b\_C; 1.

CC DR Pfam; PF00033; cytochrome\_b\_N; 1.

CC DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.

CC DR PROSITE; PS00193; CYTOCHROME\_B\_QO; 1.

CC DR KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane; Heme.

FT METAL; 83 83 IRON 1 (HEME B562 AXIAL LIGAND).

FT METAL; 97 97 IRON 2 (HEME B566 AXIAL LIGAND).

FT METAL; 182 182 IRON 2 (HEME B562 AXIAL LIGAND).

FT METAL; 196 196 IRON 1 (HEME B566 AXIAL LIGAND).

SQ SEQUENCE 379 AA; 42520 MW; F74CD03BA7C24D4 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 379;

Best local Similarity 50.0%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 31 WNFGLSLAACLAQ 44

---

RESULT 7

CYB_ZALCA	STANDARD;	PRT;	379 AA.
AC 036266;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 30-MAY-2000 (Rel. 39, Last annotation update)			
DE Cytochrome B.			
GN MTCYB OR COB OR CYTB.			
OS Zalophus californianus (California sea lion).			
OG Mitochondrion.			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Pinnipedia; Otariidae; Zalophus.			
OC NCBI_TaxID=9704;			
RN [1]			
RE SEQUENCE FROM N.A.			

---

RESULT 8

PEL_ERWCH	PEL_ERWCH	STANDARD;	PRT;	425 AA.
ID PEL_ERWCH	AC 047473; Q594121;			
DT 15-DEC-1998 (Rel. 37, Created)				
DT 15-DEC-1998 (Rel. 37, Last sequence update)				
DT 15-JUL-1999 (Rel. 38, Last annotation update)				
DE Pectate lyase 1 precursor (EC 4.2.2.2) (Pectate transeliminase).				
GN PEL_ERWCH.				
RN [1]				
RE SEQUENCE FROM N.A.				
RC STRAIN=EC16;				
RX MEDLINE=95562691; PubMed=7635842;				
RA Alfonso J.R., Ham J.H., Collmer A.;				
RT Use of intact pectate lyase 1 precursor to clone a pel gene encoding a highly alkaline, asparagine-rich pectate lyase isozyme from an Erwinia chrysanthemi EC16 mutant with deletions affecting the major pectate lyase isozymes. J. Bacteriol. 177:4553-4556(1995).				
RN [2]				
RE SEQUENCE FROM N.A.				
RC STRAIN=3937;				

RX MEDLINE=96020649; PubMed=8577252;  
 RA Lojkowska E., Masciaux C., Boccardo M., Robert-Baudouy J.,  
 RA Hugouvieux-Cotte-Pattat N.;  
 RT "Characterization of the pectL gene encoding a novel pectate lyase of  
 RT Erwinia chrysanthemi 3937.";  
 RL Mol. Microbiol. 16:1183-1195(1995).  
 CC -I - FUNCTION: PRESENTS AN ENDO-CLEAVING ACTIVITY ON POLYGLACTURONATE  
 CC OR PARTIALLY METHYLATED PECTIN.  
 CC -I - CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give  
 CC oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at  
 CC their non-reducing ends.  
 CC -I - COFACTOR: REQUIRES CALCIUM.  
 CC -I - SUBCELLULAR LOCATION: Secreted.  
 CC -I - INDUCTION: BY PECTIC CATABOLIC PRODUCTS AND AFFECTED BY GROWTH  
 CC PHASE, TEMPERATURE, IRON STARVATION, OSMOLARITY, ANAEROBOSIS,  
 CC NITROGEN STARVATION AND CATABOLITE REPRESSION.  
 CC -I - SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 9.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; XBL136; CAIA97041; 1; -.  
 DR EMBL; L42248; AAA99476; 1; -.  
 DR EMBL; XBL136; CAIA97041; 1; -.  
 KW Lyase; Signal; Calcium.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 425 PEPTIDE LYASE L.  
 FT VARIANT 49 49 S -> N (IN STRAIN 3937).  
 FT VARIANT 55 55 N -> S (IN STRAIN 3937).  
 FT VARIANT 59 59 S -> A (IN STRAIN 3937).  
 FT VARIANT 64 64 T -> S (IN STRAIN 3937).  
 FT VARIANT 70 70 A -> V (IN STRAIN 3937).  
 FT VARIANT 103 105 EG5 -> DGA (IN STRAIN 3937).  
 FT VARIANT 147 145 I -> V (IN STRAIN 3937).  
 FT VARIANT 160 160 T -> I (IN STRAIN 3937).  
 FT VARIANT 219 219 Q -> P (IN STRAIN 3937).  
 FT VARIANT 224 224 G -> V (IN STRAIN 3937).  
 FT VARIANT 261 261 S -> N (IN STRAIN 3937).  
 FT VARIANT 264 264 S -> A (IN STRAIN 3937).  
 FT VARIANT 328 328 K -> Q (IN STRAIN 3937).  
 FT VARIANT 343 348 GSATVN -> ASYIVS (IN STRAIN 3937).  
 FT VARIANT 362 362 V -> A (IN STRAIN 3937).  
 FT VARIANT 378 378 I -> V (IN STRAIN 3937).  
 FT VARIANT 390 390 A -> S (IN STRAIN 3937).  
 FT VARIANT 395 395 T -> A (IN STRAIN 3937).  
 SQ 425 AA; 45537 MW; BB2557332CFP91F7 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 425;  
 Best Local Similarity 46.7%; Pred. No. 40; Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0; OX NCBI\_TAXID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K1 / MG1655;  
 RX MEDLINE=95334362; PubMed=7610040;  
 RL Burland V.D., Blunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blunkett F.R.; Blattner F.R., Blunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA REVITIONS.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Blunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=2104935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Bimelaite E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Weich R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli 0157:H7."  
 RL Nature 409:529-533(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RIMD 050952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT 0157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -I - SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -I - SIMILARITY: BELONGS TO THE ARCD/CADB/GERRAB/LYSI FAMILY OF  
 CC PERMEASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; U14003; AA99701; 1; ALT\_FRAME.  
 DR EMBL; U14003; AA99704; 1; ALT\_FRAME.  
 DR EMBL; AB00084; AAC77076; 1; .  
 DR EMBL; AE00564; AAG95914; 1; .  
 DR EMBL; AP02568; BAB38520; 1; .  
 DR Ecogene; EG12462; YJDE.  
 DR InterPro; IPRO02293; AArel\_primeasel.  
 DR InterPro; IPRO04641; Permease.  
 DR Fram; PR00324; aa-permeases; 1.  
 DR EMBL; AA99701; 1; .  
 DR EMBL; AA99704; 1; .  
 DR TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 39 59 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 123 143 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 197 217 POTENTIAL.

RESULT 9

YJDE\_ECOLI STANDARD; PRT; 445 AA.

AC P39269; P39268; 01-FEB-1995 (Rel. 31, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical transport protein YJDE.  
 DE YJDE OR B4115 OR 25717 OR ESS5097.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.

DE (L-ESAT).  
 GN ESM16 OR ESX CCR 145 OR ML0049 OR MLCB628\_12C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1769;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP Wielies B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 SEQUENCE FROM N.A.  
 RP SROUNCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Egelmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jigels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Bairelli B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1 - SUBCELLULAR LOCATION: Secreted (BY similarity).  
 CC -1 - SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X90946; CAM62441.1; -.  
 DR EMBL; Y14967; CAAT5200.1; -.  
 DR EMBL; AU583917; CAC29557.1; -.  
 DR Lepromo; ML0049; -.  
 KW Complete proteome.  
 FT CONFLICT 55 R -> Q (IN REF. 1).  
 FT CONFLICT 90 90 M -> T (IN REF. 1).  
 SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB81 CRC64;  
 Query Match 49.4%; Score 38; DB 1; Length 95;  
 Best Local Similarity 33.3%; Pred. No. 15; Gaps 0;  
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 DB 5 WHFPLAQAVNELOG 19  
 SEQUENCE 990 AA; 11073 MW; 0664DIEFFEO38F606 CRC64;

RESULT 11

Query Match 50.6%; Score 39; DB 1; Length 990;  
 Best Local Similarity 46.7%; Pred. No. 86; Gaps 0;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 DB 927 WNTVYLERATSALRG 941

ES6\_MCLE

ID ESM6\_MCLE STANDARD; PRT; 95 AA.

AC Q50206; O33083; [1]  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 6 kDa early secretory antigenic target homolog (ESAT-6-like protein)

DE (L-ESAT).  
 GN ESM16 OR ESX CCR 145 OR ML0049 OR MLCB628\_12C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1769;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP Wielies B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 SEQUENCE FROM N.A.  
 RP SROUNCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Egelmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jigels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Bairelli B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1 - SUBCELLULAR LOCATION: Secreted (BY similarity).  
 CC -1 - SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X90946; CAM62441.1; -.  
 DR EMBL; Y14967; CAAT5200.1; -.  
 DR EMBL; AU583917; CAC29557.1; -.  
 DR Lepromo; ML0049; -.  
 KW Complete proteome.  
 FT CONFLICT 55 R -> Q (IN REF. 1).  
 FT CONFLICT 90 90 M -> T (IN REF. 1).  
 SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB81 CRC64;  
 Query Match 49.4%; Score 38; DB 1; Length 95;  
 Best Local Similarity 33.3%; Pred. No. 15; Gaps 0;  
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 DB 5 WHFPLAQAVNELOG 19  
 SEQUENCE 990 AA; 11073 MW; 0664DIEFFEO38F606 CRC64;

RESULT 12

CADB\_ECOLI

ID CADB\_ECOLI STANDARD; PRT; 444 AA.

AC P23891;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable cadaverine/lysine amidotransf

CADB\_ECOLI

OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia;  
 OC NCBI\_TaxID=562; 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9105022; PubMed=1370290;  
 RA Watson N., Dunyak D.S., Rosey E.L., Slonczewski J.L., Olson E.R.;  
 RT CADB OR B4132 OR 25735 OR EC5514.

DE (L-ESAT).  
 GN ESM16 OR ESX CCR 145 OR ML0049 OR MLCB628\_12C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1769;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP Wielies B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 SEQUENCE FROM N.A.  
 RP SROUNCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Egelmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jigels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Bairelli B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1 - SUBCELLULAR LOCATION: Secreted (BY similarity).  
 CC -1 - SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X90946; CAM62441.1; -.  
 DR EMBL; Y14967; CAAT5200.1; -.  
 DR EMBL; AU583917; CAC29557.1; -.  
 DR Lepromo; ML0049; -.  
 KW Complete proteome.  
 FT CONFLICT 55 R -> Q (IN REF. 1).  
 FT CONFLICT 90 90 M -> T (IN REF. 1).  
 SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB81 CRC64;  
 Query Match 49.4%; Score 38; DB 1; Length 95;  
 Best Local Similarity 33.3%; Pred. No. 15; Gaps 0;  
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 DB 5 WHFPLAQAVNELOG 19  
 SEQUENCE 990 AA; 11073 MW; 0664DIEFFEO38F606 CRC64;

RESULT 12

CADB\_ECOLI

ID CADB\_ECOLI STANDARD; PRT; 444 AA.

AC P23891;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable cadaverine/lysine amidotransf

CADB\_ECOLI

OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia;  
 OC NCBI\_TaxID=562; 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9105022; PubMed=1370290;  
 RA Watson N., Dunyak D.S., Rosey E.L., Slonczewski J.L., Olson E.R.;  
 RT CADB OR B4132 OR 25735 OR EC5514.

DE (L-ESAT).  
 GN ESM16 OR ESX CCR 145 OR ML0049 OR MLCB628\_12C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1769;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP Wielies B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 SEQUENCE FROM N.A.  
 RP SROUNCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Egelmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jigels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Bairelli B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 CC -1 - SUBCELLULAR LOCATION: Secreted (BY similarity).  
 CC -1 - SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X90946; CAM62441.1; -.  
 DR EMBL; Y14967; CAAT5200.1; -.  
 DR EMBL; AU583917; CAC29557.1; -.  
 DR Lepromo; ML0049; -.  
 KW Complete proteome.  
 FT CONFLICT 55 R -> Q (IN REF. 1).  
 FT CONFLICT 90 90 M -> T (IN REF. 1).  
 SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB81 CRC64;  
 Query Match 49.4%; Score 38; DB 1; Length 95;  
 Best Local Similarity 33.3%; Pred. No. 15; Gaps 0;  
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 DB 5 WHFPLAQAVNELOG 19  
 SEQUENCE 990 AA; 11073 MW; 0664DIEFFEO38F606 CRC64;



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_licensee@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_licensee@isb-sib.ch)).

CC EMBL; 051269; Ac51202.1; -  
 CC HSSP; P35222; I633J.  
 CC Genew; HGNC:7128; ARVCF.  
 CC MIM; 602269; -  
 CC Interpro; IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 6.  
 DR SMART; SM00185; ARM\_4.  
 DR PROSITE; PSS0176; ARM\_REPEAT; 3.  
 DR KW Alternative splicing; Coiled coil; Cell adhesion; Cytoskeleton; structural protein; Repeat; Coiled coil; KW DOMAIN 8 46 COILED COIL (POTENTIAL).  
 FT DOMAIN 607 623 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT REPEAT 608 611 POLY-ARG.  
 FT REPEAT 348 387 ARM 1.  
 FT REPEAT 390 429 ARM 2.  
 FT REPEAT 433 467 ARM 3.  
 FT REPEAT 468 508 ARM 4.  
 FT REPEAT 526 565 ARM 5.  
 FT REPEAT 575 622 ARM 6.  
 FT REPEAT 646 686 ARM 7.  
 FT REPEAT 699 738 ARM 8.  
 FT REPEAT 739 781 ARM 9.  
 FT REPEAT 782 826 ARM 10.  
 FT VARSPlic 1 69 MEDCNVRASILASVKQEAEARRERQERRHVALQL ERQQPGNNSGNGSGQPLPMWQQLVL -> MPAELR  
 FT SEQUENCE 962 AA; 104641 MW; 74A1814A022FF2B1 CRC64;  
 SQ

RESULT 14  
 ARVC\_HUMAN STANDARD; PRT; 962 AA.  
 ID ARVC\_HUMAN STANDARD; PRT; 962 AA.  
 AC 000192; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome.  
 GN ARVCF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metacara; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1] RZ MEDLINE=9721559; PubMed=912645;  
 RA Sirotnik H., O'donnell H., Dassgupta R., Halford S., St Jore B.,  
 RA Puech A., Parimoo S., Morrow B., Skoultschi A., Weissman S.,  
 RA Scambler P., Kucherlapati R.;  
 RA "Identification of a new human catenin gene family member (ARVCF) from the region deleted in velo-cardio-facial syndrome.";  
 RA Genomics 41:75-83(1997).  
 CC -I- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS JUNCTIONS.  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM AND A LONG FORM (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPlicing.  
 CC -I- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING HEART, BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.  
 CC -I- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS HEMIZYGOUS IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS. THIS HEMIZYGOSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM PHENOTYPES, INCLUDING CONONTRUNCAL HEART DEFECTS, CLFRT PALATE AND FACIAL DYSMORPHOLOGY.  
 CC -I- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -I- SIMILARITY: CONTAINS 10 ARM REPEATS.

Query Match Score 38; DB 1; Length 923;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 NFAGTCRAASATQ 15  
 DB 8311 DFAGVADATLLEG 844  
 FT SEQUENCE 962 AA; 104641 MW; 74A1814A022FF2B1 CRC64;  
 SQ

RESULT 15  
 ARVC\_MOUSE STANDARD; PRT; 969 AA.  
 ID ARVC\_MOUSE STANDARD; PRT; 969 AA.  
 AC P08203; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome  
 DE homolog (Fragment).  
 GN ARVCF.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metacara; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1] RZ MEDLINE=20512094; PubMed=11058098;  
 RA Kaufmann U., Zuppinger C., Waibler Z., Rudiger M., Urbich C.,  
 RA Martin B., Jockusch B.M., Eppenberger H., Starzinski-Powitz A.;  
 RT The armadillo repeat region targets ARVCF to cadherin-based cellular junctions.";  
 RT RL J. Cell Sci. 113:4121-4135(2000).  
 CC -I- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS JUNCTIONS (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -I- SIMILARITY: CONTAINS 10 ARM REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_licensee@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_licensee@isb-sib.ch)).

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_licensee@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_licensee@isb-sib.ch)).  
 CC EMBL; 051269; Ac51202.1; -  
 CC HSSP; P35222; I633J.  
 CC Genew; HGNC:7128; ARVCF.  
 CC MIM; 602269; -  
 CC Interpro; IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 6.  
 DR SMART; SM00185; ARM\_4.  
 DR PROSITE; PSS0176; ARM\_REPEAT; 3.  
 DR KW Alternative splicing; Coiled coil; Cell adhesion; Cytoskeleton; structural protein; Repeat; Coiled coil; KW DOMAIN 8 46 COILED COIL (POTENTIAL).  
 FT DOMAIN 607 623 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT REPEAT 608 611 POLY-ARG.  
 FT REPEAT 348 387 ARM 1.  
 FT REPEAT 390 429 ARM 2.  
 FT REPEAT 433 467 ARM 3.  
 FT REPEAT 468 508 ARM 4.  
 FT REPEAT 526 565 ARM 5.  
 FT REPEAT 575 622 ARM 6.  
 FT REPEAT 646 686 ARM 7.  
 FT REPEAT 699 738 ARM 8.  
 FT REPEAT 739 781 ARM 9.  
 FT REPEAT 782 826 ARM 10.  
 FT VARSPlic 1 69 MEDCNVRASILASVKQEAEARRERQERRHVALQL ERQQPGNNSGNGSGQPLPMWQQLVL -> MPAELR  
 FT SEQUENCE 962 AA; 104641 MW; 74A1814A022FF2B1 CRC64;  
 SQ

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to license@isb-sib.ch).

DR EMBL: AD243418; -; NOT\_ANNOTATED\_CDS.

DR HSSP: P15222; 1G3J.

DR MGD: MG1:1196220; Arvcf.

DR InterPro: IPR000225; Armadillo.

DR Pfam: PF00514; Armadillo\_seg; 6.

DR SMART: SM0015; ARM; 4.

DR PROSITE: PS50176; ARM\_REPEAT; 3.

KW Cell adhesion; Cytoskeleton; structural protein; Repeat; Coiled coil.

FT NON\_TER 1 1 COILED\_COIL\_(POTENTIAL).

FT DOMAIN <1 39 ARM 1.

FT REPEAT 328 376 ARM 2.

FT REPEAT 377 418 ARM 3.

FT REPEAT 419 462 ARM 3.

FT REPEAT 463 503 ARM 4.

FT REPEAT 521 560 ARM 5.

FT REPEAT 570 616 ARM 6.

FT REPEAT 634 674 ARM 7.

FT REPEAT 675 720 ARM 8.

FT REPEAT 721 769 ARM 9.

FT REPEAT 770 814 ARM 10.

SQ 969 AA: 103378 MW: 91130069E484C5B5 CRC64:

Query Match 49.4%; Score 38; DB 1; Length 969;  
Best Local Similarity 61.5%; Pred. No. 1.2e+02; Mismatches  
Matches 8; Conservative 2; Gaps 0; Gaps 0;

Qy 2 NFAGCIEAANSQ 14  
|| : ||| | : |  
Db 664 NFTLEAAAGALQ 676

Search completed: July 3, 2003, 14:05:52  
Job time : 7.5 secs

Run on:	July 3, 2003, 14:03:50 ;	(without alignments)	Total number of hits satisfying chosen parameters:	283224
Minimum DB seq length:	0	Maximum DB seq length:	200000000	Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries
Database :	PIR_73:*	1: pir1:*	No. 1 is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	2
Scoring table:	BLOSSUM62	2: pir2:*		3
Searched:	Gappp 10.0 , Gapext 0.5	3: pir3:*		4
	283224 seqs, 96134422 residues	4: pir4:*		45
Pred				
No. 1 is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.				
SUMMARIES				
Result No.	Score	Query Match Length	DB ID	Description
1	77	100	0	A70803 early secretory antigen target esat5 - Mycobacterium tuberculosis (strain H37RV)
2	47	61.0	714	G66844 C;Species: Mycoacterium tuberculosis
3	44	57.1	2	C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999 C;Accession: A70803; S49174 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, Connor, R.; Davies, R.; Devlin, K.; Feilwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
4	43	55.8	476	A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A;Reference number: A70500; MUID: 98295987; PMID: 9634230
5	43	55.8	2	A;Accession: A70803 A;Status: nucleic acid sequence not shown; translation not shown A;Molecule type: DNA A;Residues: 1-95 <COL> A;Cross-references: GB:AL022120; GB:AL123456; NID:93261558; PID:CAA17967.1; PID:e126 A;Experimental source: strain H37Rv R;Soerensen, A.L.; Nagai, S.; Rouen, G.; Andersen, P.; Andersen, A.B. submitted to the EMBL Data Library, June 1994 A;Reference number: S49174 A;Accession: S49174 A;Molecule type: DNA A;Residues: 1-13, R, 15-22, 'S, 24-95 <SOE> A;Cross-references: EMBL:X79562; NID:9531708 A;Genetics: A;Gene: esat6
6	42.5	55.2	141	H83229 conserved hypothetical protein
7	42	54.5	369	E86554 Fe-S oxidoreductas
8	42	54.5	369	G72069 conserved hypothetical protein
9	41	53.2	384	A69622 ferrichrome ABC transporter
10	41	53.2	445	D82349 probable cadaverin ABC transporter
11	41	53.2	540	F86745 glyceraldehyde-3-p
12	40	51.9	294	BH3305 conserved hypothetical protein
13	40	51.9	323	F8979
14	40	51.9	330	AD2675
15	40	51.9	323	R69020
16	40	51.9	493	F95940 probable xanthine
17	40	51.9	536	B43439
18	40	51.9	761	T51912 hypothetical prote
19	39	50.6	101	F71052 hypothetical prote
20	39	50.6	287	B71063 hypothetical prote
21	39	50.6	379	S88457 ubiquinol-cytochrome
22	39	50.6	379	ubiquinol-cytochrome
23	39	50.6	445	B85221 probable amino acid
24	39	50.6	445	A91266 probable amino acid
25	39	50.6	445	F86106 probable amino acid
26	39	50.6	445	AF1022 tyrosine amino acid
27	39	50.6	448	T05591 tyrosine transamin
28	39	50.6	715	B83518 palmitoyl-CoA synth
29	39	50.6	745	hypothetical protein
30	39	50.6	990	transposase - EcBC
31	38	49.4	95	early secretory antigen target esat5 - Mycobacterium tuberculosis (strain H37RV)
32	38	49.4	98	conserved hypothetical protein
33	38	49.4	195	hypothetical protein
34	38	49.4	219	hypothetical protein
35	38	49.4	303	hypothetical protein
36	38	49.4	313	hypothetical protein
37	38	49.4	313	hypothetical protein
38	38	49.4	399	hypothetical protein
39	38	49.4	422	tyrosine transamin
40	38	49.4	425	pectate lyase (EC
41	38	49.4	443	probable cadaverine
42	38	49.4	444	lysine/cadaverine
43	38	49.4	444	transport protein
44	38	49.4	444	probable amino acid
45	38	49.4	444	AE0147
ALIGNMENTS				
RESULT 1				
A70803				
early secretory antigen target esat5 - Mycobacterium tuberculosis (strain H37RV)				
C;Species: Mycoacterium tuberculosis				
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999				
C;Accession: A70803; S49174				
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, Connor, R.; Davies, R.; Devlin, K.; Feilwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998				
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome				
A;Reference number: A70500; MUID: 98295987; PMID: 9634230				
A;Status: nucleic acid sequence not shown; translation not shown				
A;Molecule type: DNA				
A;Residues: 1-95 <COL>				
A;Cross-references: GB:AL022120; GB:AL123456; NID:93261558; PID:CAA17967.1; PID:e126				
A;Experimental source: strain H37Rv				
R;Soerensen, A.L.; Nagai, S.; Rouen, G.; Andersen, P.; Andersen, A.B. submitted to the EMBL Data Library, June 1994				
A;Reference number: S49174				
A;Accession: S49174				
A;Molecule type: DNA				
A;Residues: 1-13, R, 15-22, 'S, 24-95 <SOE>				
A;Cross-references: EMBL:X79562; NID:9531708				
A;Genetics: A;Gene: esat6				
Query Match				
Best Local Similarity				
Matches				
15; Conservative				
QY				
1 WNFAGIEAAASATQG 15				
DB				
6 WNFAGIEAAASATQG 20				
RESULT 2				
G86844				
hypothetical protein glnP [imported] - Lactococcus lactis subsp. lactis (strain IL140)				
C;Species: Lactococcus lactis subsp. lactis				
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001				
C;Accession: G86844				
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh				
Genome Res. 11, 731-753, 2001				
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis				
A;Reference number: A86625; MUID: 21235186; PMID: 11337471				
A;Accession: G86844				
A;Status: preliminary				
A;Molecule type: DNA				
A;Residues: 1-714 <STO>				

A;Cross-references: GB:AE005176; PID:912724781; PIDN:AAK05857.1; GSPDB:GN00146  
A;Experimental source: strain IL403  
C;Genetics:  
A;Gene: glnP

Query Match Similarity 61.0%; Score 47; DB 2; Length 714;  
Best Local Similarity 50.0%; Pred. No. 4.5%;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WNFAGIEAAA10 14  
Db 303 WNFAGFOAADSVO 316

RESULT 3

T50707 urease accessory protein D [imported] - Rhodobacter sphaeroides  
C;Species: Rhodobacter sphaeroides  
C;Accession: T50707  
R;Choudhary, M.; Kaplan, S.  
Nucleic Acids Res. 28, 862-867, 2000  
A;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides  
A;Reference number: Z25222; MUID:20115911; PMID:10641776  
A;Accession: T50707  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-225 <CHO>  
A;Cross-references: EMBL:AF195122; PIDN:AAF24251.1  
A;Experimental source: strain 2.4.1  
C;Genetics:  
A;Gene: ureD

Query Match Similarity 57.1%; Score 44; DB 2; Length 225;  
Best Local Similarity 75.0%; Pred. No. 4.5%;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AGIERRASATQ 15  
Db 178 AGVEARASAfdG 189

RESULT 4

S71360 noea protein - Rhizobium meliloti  
C;Species: Rhizobium meliloti  
C;Accession: S71360 #sequence\_revision 13-Mar-1997 #text\_change 04-Mar-2000  
R;Ridoulet, M.; Lorette, G.; Maillet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosenberg, M.; submitted to the EMBL Data Library, May 1995  
A;Reference number: S71360  
A;Accession: S71360  
A;Molecule type: DNA  
A;Residues: 1-476 <ARD>  
A;Cross-references: EMBL:U26430; NID:91326068; PIDN:MAC44091.1; PID:91326070  
A;Experimental source: strain RCF2011  
R;Ardourel, M.; Lorette, G.; Maillet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosenberg, M.; Microbiol. 17, 687-699, 1995  
A;Title: In Rhizobium meliloti, the operon associated with the nod box n5 comprises nodD, a reference number: S71357; MUID:96111489; PMID:8801423  
A;Accession: S71358  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 304-339 <ARW>  
A;Cross-references: EMBL:U26430  
A;Experimental source: strain RCF2011 (=SU47)  
C;Genetics:  
A;Gene: noea  
C;Superfamily: Rhizobium meliloti noea protein

Query Match Similarity 55.8%; Score 43; DB 2; Length 476;  
Best Local Similarity 70.0%; Pred. No. 15%;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WNFAGIEAAA 10  
Db 101 WSFAGLKAA 110

RESULT 5

H9313 H9313 Noea host specific nodulation protein [imported] - Sinorhizobium meliloti (strain 102  
C;Accession: H9313  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Species: Sinorhizobium meliloti  
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.; Proc. Natl. Acad. Sci. U.S.A. 98, 9833-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli  
A;Reference number: A95262; MUID:21396509; PMID:11481432  
A;Accession: H9313  
A;Molecule type: DNA  
A;Strains: preliminary  
A;Cross-references: GB:AE006469; PIDN:AAK55074.1; PID:914523508; GSPDB:SN00165  
A;Experimental source: strain 1021, megaplasmid pSYMP  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Fedderspiel, N.A.; Fisher, R.; L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
heault, P.; Vandenberg, M.; Vorholter, F.J.; Weindner, S.; Wells, D.H.; Wong, K.; Yeh,  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21360234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: noea

Query Match Similarity 55.2%; Score 43; DB 2; Length 476;  
Best Local Similarity 70.9%; Pred. No. 5.1%;  
Matches 9; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

Query Match Similarity 55.8%; Score 43; DB 2; Length 141;  
Best Local Similarity 40.9%; Pred. No. 5.1%;  
Matches 9; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

Query Match 55.8%; Score 43; DB 2; Length 476;

Oy 1 WN-----FAGIEAAASAIQG 15

Db ||| 1:||||:||| 119 WNPRLRHLALGGVEAAKIVQG 140

## RESULT 7

E6554

Fes-S oxidoreductase [imported] - *Chlamydophila pneumoniae* (strain J138)C;Species: *Chlamydophila pneumoniae*, *Chlamydia pneumoniae*

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: E6554

R; Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is-

Nucleic Acids Res. 28, 2311-2314, 2000

A;Title: Comparison of whole genome sequences of *chlamydia pneumoniae* J138.

A;Reference number: A86491; MUID:20330349; PMID:10871362

A;Accession: E6554

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-369 &lt;SPO&gt;

A;Cross-references: GB:BA000008; NID:98978883; PIDN:BAA98719.1; GSPDB:GN00142

A;Experimental source: strain J138

C;Genetics:

C;Gene: CP0513

C;Superfamily: hypothetical protein AF0390

Query Match 54.5%; Score 42; DB 2; Length 369;

Best Local Similarity 77.8%; Pred. No. 17;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 WNFAIGEAA 9

Db 305 WNYLGIETAA 313

## RESULT 8

G72059

Conserved hypothetical protein CP0240 [imported] - *Chlamydophila pneumoniae* (strain CWI)

N;Alternate names: fe-s oxidoreductase

C;Species: *Chlamydophila pneumoniae*, *Chlamydia pneumoniae*

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C;Accession: GT2199

R; Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.

A;Reference number: A72000; MUID:99206606; PMID:10192388

A;Accession: GT2069

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-369 &lt;ARN&gt;

A;Cross-references: GB:AE001636; GB:AE001363; NID:94376794; PIDN:AAD18653.1; PID:9437680

A;Experimental source: strain CWL029

R; Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kotonya, J.; McClarty, G.; Salzberg, S. Nucleic Acids Res. 28, 1397-1405, 2000

A;Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: F81599

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-369 &lt;RNA&gt;

A;Cross-references: GB:AE002184; GB:AE002161; NID:97189156; PIDN:AAF38105.1; PID:9718916

A;Experimental source: strain AR39, HL cells

C;Genetics:

C;Gene: CP0513; CP0340

C;Superfamily: hypothetical protein AF0390

## RESULT 9

A69622

ferrichrome ABC transporter (permease) fhub - *Bacillus subtilis*C;Species: *Bacillus subtilis*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C;Accession: A69622

R; Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bar

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Einlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferari,

A.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

Tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hollsappel, S.; Hullo, M.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Hardino

A.; Authors: Laufer, J.; Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y.; Ogawa, K.; Ogawa, K.; Oudega, B.; Park, S.H.; Parco, V.; Poli, T.M.; Porte

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sato, T.; Scanl

A.; Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiuchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tomato, A.; Uchida

T.; Winters, P.; Wipat, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A.; Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A.; Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: A69622

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-384 &lt;KON&gt;

A;Cross-references: GB:Z99121; GB:AL009126; NID:92635827; PIDN:CAB15337.1; PID:926358

A;Experimental source: strain 168

C;Genetics:

C;Gene: fhub

C;Superfamily: vitamin B12 transport protein btuc

## RESULT 10

D82342

probable cadaverine/Lysine antiporter CadB VC0280 [imported] - *vibrio cholerae* (strainC;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: DB2342

R; Heidelberg, F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

J.; Brinkac, L.M.; Charon, D.; Eraut, M.D.; Fraser, C.M.; Perna, N.T.; Mekalanos, J.J.; Ventre, J.C.; Fraser, C.M.

I.; R.R.; Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: DB2342

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-445 &lt;HEI&gt;

A;Cross-references: GB:AE004116; GB:AE003852; NID:9965687; PIDN:AAF33455.1; GSPDB:GN

A;Experimental source: serogroup O1; strain NI6961; biotype El Tor

C;Genetics:

C;Gene: VC0280

A;Map position: 1

C;Superfamily: L-Lysine transport protein

## RESULT 11

Oy 1 WNFAIGEAA 9

Query Match 54.5%; Score 42; DB 2; Length 369;

Best Local Similarity 77.8%; Pred. No. 17;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 WNYLGIETAA 10

||:||||:|||

Db 199 WSFVGVESAA 208

RESULT 11  
FB6745  
ABC transporter permease protein YJJD [imported] - Lactococcus lactis subsp. lactis (strain C) Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C;Accession: R86745  
R.Bolotin, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrling, S.; Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis subsp. lactis (strain C) Species: Lactococcus lactis subsp. lactis  
A;Reference number: AB6625; MUID:21235186; PMID:11337471  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-540 <STO>  
A;Cross-references: GB:AE005176; PIDN:912723906; PIDN:AAK05064\_1; GSPDB:GN00146  
A;Experimental source: strain IL1403  
A;Gene: YJJD

Query Match 53.2%; Score 41; DB 2; Length 540;  
Best Local Similarity 69.2%; Pred. No. 37; Mismatches 8; Indels 3; Gaps 0;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 FAGIEAAASAIQG 15  
Db 339 FVGENTAASAIRG 351

RESULT 12  
I41220  
9-glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - Escherichia coli Species: Escherichia blattae  
C;Date: 04-Sep-1997 #sequence\_revision 07-Nov-1997 #text\_change 03-Jun-2002  
C;Accession: I41220; I41223; I41226  
C;Lawrence, J.G.; Ochman, H.; Hartl, D.L.  
J. Gen. Microbiol. 137, 1911-1921, 1991  
A;Title: Molecular and evolutionary relationships among enteric bacteria.  
A;Reference number: I40701; MUID:92065252; PMID:1955870  
A;Accession: I41220  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-294 <RES>  
A;Cross-references: GB:M63358; NID:9146076; PIDN:AAA23837.1; PID:9146077  
A;Accession: I41223  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-294 <RES>  
A;Cross-references: GB:M63359; NID:9146108; PIDN:AAA23852.1; PID:9146109  
A;Experimental source: ATCC 33429  
A;Accession: I41226  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-294 <RES>  
A;Cross-references: GB:M63360; NID:9146114; PIDN:AAA23855.1; PID:9146115  
A;Experimental source: ATCC 33430  
C;Superfamily: glyceraldehyde-3-phosphate dehydrogenase  
C;Keywords: oxidoreductase

Query Match 53.2%; Score 41; DB 2; Length 540;  
Best Local Similarity 69.2%; Pred. No. 37; Mismatches 8; Indels 3; Gaps 0;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 FAGIEAAASAIQG 15  
Db 339 FVGENTAASAIRG 351

RESULT 14  
AD2675  
conserved hypothetical protein Atu0802 [imported] - Agrobacterium tumefaciens (strain C) Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C;Accession: AD2675  
R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, E.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, R.; Gilpin, P.; Romero, P.; Zhang, S.; Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, R.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; PMID:11743193  
A;Accession: AD2675  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-323 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AM41818.1; PID:917739175; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
A;Gene: Atu0802  
A;Map position: circular chromosome

Query Match 51.9%; Score 40; DB 2; Length 323;  
Best Local Similarity 61.5%; Pred. No. 32; Mismatches 8; Indels 2; Gaps 0;

Qy 3 FAGIEAAASAIQG 15  
Db 11 FRAGISCATALAG 23

RESULT 15  
H69020  
probable Na+/Ca2+, K+-exchanging protein - Methanobacterium thermoautotrophicum (strain C) Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C;Accession: H69020  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicinie, R.; Wang, Y.; Wierzowski, J.; Gibson, R.; Jiwanji, S.; Church, G.M.; Daniels, C.J.; Mac, J.; Rice, P.; Noelting, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu  
A;Reference number: A69000; MUID:9037514; PMID:9371463  
A;Accession: H69020  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

RESULT 13  
B97457  
BH3305 conserved hypothetical [imported] - Agrobacterium tumefaciens (strain C58, Cereon

Thu Jul 3 14:50:04 2003

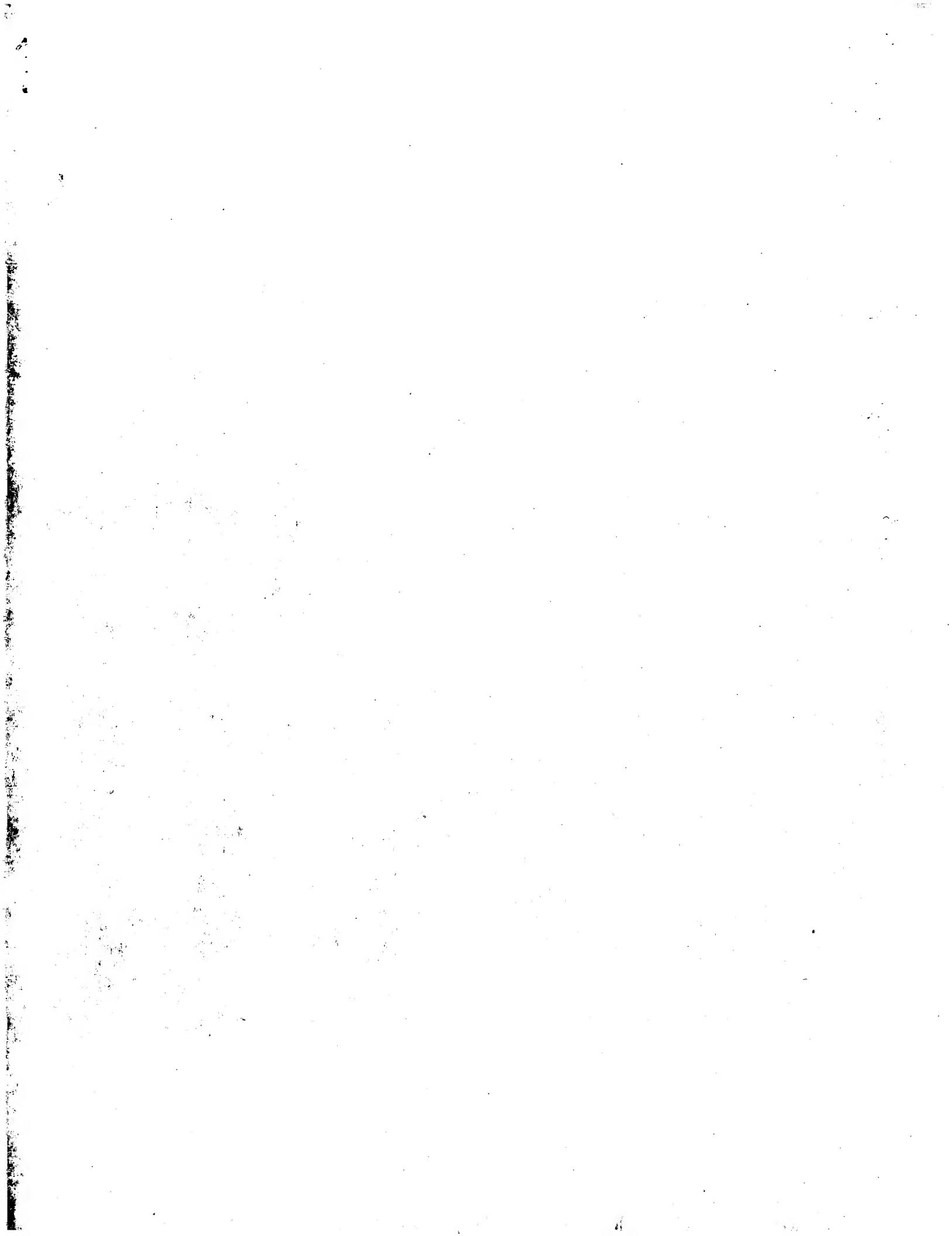
us-09-830-839-6.rpr

Page 5

A; Molecule type: DNA  
A; Residues: 1-330 <WTH>  
A; Cross-references: GB:AE000885; GB:AE000666; NID:92622256; PID:AAB85644.1; PID:9262226  
A; Experimental source: strain Delta H  
C; Genetics:  
A; Gene: MTU1155  
A; Start codon: TTG

Query Match 51.9%; Score 40; DB 2; Length 330;  
Best Local Similarity 46.7%; Pred. No. 33;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 WNFAGIEAAASIQG 15  
|| ||| :| :|  
Db 95 WNTAGILGISSVWAG 109

Search completed: July 3, 2003, 14:07:19  
Job time : 13.5 secs



Gercore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 3, 2003, 14:00:24 ; Search time 30 Seconds  
(without alignments)  
66.625 Million cell updates/sec

Title: US-09-830-839-6  
Perfect score: 77  
Sequence: 1 WNFAGGIEAAASATQG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI991.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI2001.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AAI2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	15	21	AAV9495
2	77	100.0	15	23	AAV17433
3	77	100.0	20	22	AB1225
4	77	100.0	51	18	AAW32466
5	77	100.0	51	18	AAW32339
6	77	100.0	51	19	AAW81701
7	77	100.0	51	19	AAW64334
8	77	100.0	51	20	AAV3913
9	77	100.0	51	20	AAV3898
10	100.0	51	22		AAU01897

## ALIGNMENTS

RESULT 1	AAV94595	standard; peptide; 15 AA.
ID	AAV94595	
XX		
AC	AAV94595;	
XX		
DT	28-NOV-2000 (first entry)	
XX		
DE	Mycobacterium tuberculosis ESAT-6 peptide ESG6.	
XX		
KW	Tuberculosis; infection diagnosis; ESAT-6.	
XX		
OS	Mycobacterium tuberculosis.	
XX		
PN	W0200026248-A2.	
XX		
PD	11-MAY-2000.	
XX		
PF	03-NOV-1999; 99WO-GB03635.	
XX		
PR	04-NOV-1998; 98GB-0024213.	
PR	04-NOV-1998; 98US-0107004.	
XX		
PA	(ISSIS-) ISIS INNOVATION LTD.	
XX		
PI	Lalvani A, Pathan AA;	
XX		
DR	WPI; 2000-365579/31.	
XX		
PT	Novel method of diagnosing infection, or exposure of a host, to a microbe comprising contacting T cells from the host with ESAT-6 derived peptides.	
PT		
XX		

PS Claim 1; Page 3; 33pp; English.

XX

CC The present sequence is the peptide Es5 derived from the Mycobacterium tuberculosis ESAT-6 gene. This sequence is one of eleven peptides derived from the ESAT-6 gene (see AAY4590 to AAY4600). The peptides are recognised, to varying degrees, by the T cells of patients with tuberculosis. When the peptides are combined together as a panel they provide a high specificity and sensitivity diagnostic test for M.tuberculosis infection. This test has the advantage that it does not give false positives for patients who have been vaccinated with BCG.

CC

CC SQ Sequence 15 AA;

CC Query Match 100.0%; Score 77; DB 21; Length 15;  
CC Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 1 WNFAGIEAANASA1QG 15  
CC 1 WNFAGIEAANASA1QG 15

Db

RESULT 2

ID AAO17433

XX AAO17433 standard; peptide; 15 AA.

AC AAO17433;

XX DT 11-JUL-2002 (first entry)

XX DE Early secreted antigenic target 6 T cell epitope #2.

XX Tuberculosis; TB; vaccination; vaccine; CD4+ T cell immune response;  
KW poxvirus vector; HIV; malaria; Helicobacter pylori; influenza;  
KW hepatitis; viral infection; leprosy; protozoan parasite; cancer;  
KW tuberculostatic; anti-HIV; protozoacide; antibacterial; virucide;  
KW hepatotropic; anti-inflammatory; antileprotic; cytostatic; epitope.

OS Mycobacterium tuberculosis.

XX OS WO200221224-A2.

PN

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-US08906.

XX PR 20-MAR-2000; 2000US-190834P.

XX PA (UYBR-) UNIV BROWN RES FOUND.

XX PI Degroot AS;

XX DR WPI; 2001-616401/71.

XX PT New vaccine for immunizing a mammalian subject, preferably humans,  
XX against infection caused by Mycobacterium tuberculosis.

XX PS Disclosure; Fig 4; 42pp; English.

XX CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mtb immune response by raising anti-Mtb antibody in a mammalian subject preferably human. They are used for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide.

XX SQ Sequence 20 AA;

CC Query Match 100.0%; Score 77; DB 22; Length 20;  
CC Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 1 WNFAGIEAANASA1QG 15  
CC 1 WNFAGIEAANASA1QG 15

Db

RESULT 3

ID AAE12275

XX AAE12275 standard; peptide; 20 AA.

XX AC AAE12275;

XX DT 18-DEC-2001 (first entry)

XX DE Mycobacterium tuberculosis (Mtb) peptide #60.

XX KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;  
KW infection; anti-Mtb immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO20170774-A2.

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-US08906.

XX PR 20-MAR-2000; 2000US-190834P.

XX PA (UYBR-) UNIV BROWN RES FOUND.

XX PI Degroot AS;

XX DR WPI; 2001-616401/71.

XX PT New vaccine for immunizing a mammalian subject, preferably humans,  
XX against infection caused by Mycobacterium tuberculosis.

XX PS Disclosure; Fig 4; 42pp; English.

XX CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mtb immune response by raising anti-Mtb antibody in a mammalian subject preferably human. They are used for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide.

XX SQ Sequence 20 AA;

CC Query Match 100.0%; Score 77; DB 22; Length 20;  
CC Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC QY 1 WNFAGIEAANASA1QG 15  
CC 1 WNFAGIEAANASA1QG 15

Db

RESULT 4

ID AAN32465

XX AAN32466 standard; Protein; 51 AA.

AC AAW32466;  
 XX  
 DT 09-JAN-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen ESAT-6.  
 XX  
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M. tuberculosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9709428-A2.  
 XX  
 PD 13-MAR-1997.  
 XX  
 PF 30-AUG-1996; 96WO-US14675.  
 XX  
 PR 12-JUL-1996; 96US-0680573.  
 PR 01-SEP-1995; 95US-0523435.  
 PR 22-SEP-1995; 95US-0532136.  
 PR 22-MAR-1996; 96US-0620280.  
 PR 05-JUN-1996; 96US-0658800.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PT 13-MAR-1997.  
 XX  
 PP 30-AUG-1996; 96WO-US14674.  
 XX  
 PR 12-JUN-1996; 96US-0680574.  
 PR 01-SEP-1995; 95US-0523426.  
 PR 22-SEP-1995; 95US-053364.  
 PR 22-MAR-1996; 96US-0620874.  
 PR 05-JUN-1996; 96US-0659683.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PT Campos neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 DR WPI; 1997-192904-17.  
 XX  
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 - useful for diagnosis of M. tuberculosis infection  
 XX  
 PS Claim 43; Page 147; 190pp; English.  
 XX  
 CC A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a specifically claimed  
 CC sequence of the ESAT-6 M. tuberculosis antigen. The immunogenic  
 CC polypeptide can be used to diagnose M. tuberculosis infection by forming  
 CC complexes with specific antibodies in the sample. Fragments of DNA  
 CC encoding the immunogenic polypeptide can be used as diagnostic primers  
 CC or probes and agents that bind to the antigen, especially monoclonal  
 CC antibodies or equivalent polyclonal antibodies, are also used for  
 CC diagnosis.  
 XX  
 SQ Sequence 51 AA;  
 CC Query Match 100.0%; Score 77; DB 18; Length 51;  
 CC Best Local Similarity 100.0%; Pred. No. 5.2e-06;  
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC QY 1 WNFAGIEAAASAIQG 15  
 CC ||||||| ||||| |||||  
 CC 6 WNFAGIEAAASAIQG 20  
 XX  
 RESULT 5  
 AAW32339  
 ID AAW32339 standard; Protein; 51 AA.  
 XX  
 AC AAW32339;  
 XX  
 DT 13-JAN-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen ESAT-6.  
 XX  
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M. tuberculosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9816546-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PR 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX WPI: 1998-261042/23.  
 DR N-PDB; AAV64501.

XX Immunogenic *Mycobacterium tuberculosis* polypeptide(s) and DNA - used  
 PT to develop products for the detection of *M. tuberculosis* infection  
 and for diagnosis, treatment and prevention of tuberculosis  
 XX Disclosure; Page 126; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
 CC *Mycobacterium tuberculosis* (Mr) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against *M. tuberculosis* infection or may  
 CC be used for the diagnosis of tuberculosis.

XX Sequence 51 AA;

Query Match	Score	DB	Length
Best Local Similarity	100.0%	19	51
Matches	15;	Conservative	0;
Qy	1 WNFAGIEAASAIAQG 15		
Db	6 WNFAGIEAASAIAQG 20		

RESULT 7  
 AAW64334  
 ID AAW64334 standard; Protein; 51 AA.

XX AAW64334;  
 AC AAW64334;  
 XX 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen ESAT-6.

XX Tuberculosis; infection; diagnosis; antigen; ESAT-6.

XX Mycobacterium tuberculosis.

OS WO9816645-A2.

XX PD 23-APR-1998.

PP 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-081811.

PR 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

PA Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX WPI: 1998-527409/44.  
 DR N-PDB; AAV64503.

XX New antigens from *Mycobacterium tuberculosis* useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX Disclosure; Page 120; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a *Mycobacterium tuberculosis* antigen (Ag). Also described  
 CC are vaccines and fusion protein containing *M. tuberculosis* Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC *M. tuberculosis* and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC *tuberculosis* immune subtypes. AAV64503 and AAV39083 to  
 CC AAV3925 are used in the exemplification of the present invention.

XX Sequence 51 AA;

Query Match	Score	DB	Length
Best Local Similarity	100.0%	20	51
Matches	15;	Conservative	0;
Qy	1 WNFAGIEAASAIAQG 15		
Db	6 WNFAGIEAASAIAQG 20		

RESULT 8  
 AAY39131  
 ID AAY39131 standard; Protein; 51 AA.

XX AAY39131;  
 AC AAY39131;  
 XX 05-NOV-1999 (first entry)

DE M. tuberculosis ESAT-6 amino acid sequence.

XX Mycobacterium tuberculosis; *M. tuberculosis*; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN WO942076-A2.

XX PD 26-AUG-1999.

PP 17-FEB-1999; 99WO-US03268.

XX PR 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX DR N-PDB; AAV64503.

XX New isolated *Mycobacterium tuberculosis* polypeptides and DNA - used  
 PT to develop products for the detection of *M. tuberculosis* infection  
 PT and diagnosis of tuberculosis.

XX Disclosure; Page 130; 250pp; English.

XX This polypeptide comprises an antigenic portion of *Mycobacterium*  
 CC tuberculosis antigen ESAT-16. The invention relates to  
 CC compositions and methods for diagnosing tuberculosis. It provides  
 CC polypeptides (see AAW64291-64379) comprising an antigenic portion  
 CC of a soluble *M. tuberculosis* antigen, or an immunogenic portion of





CC M. tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used  
 CC to diagnose ongoing/previous sensitisation with these bacteria by  
 CC detecting cytokine release when contacting blood samples with the  
 CC polypeptide. The bioreactive polypeptide may be used in diagnostic  
 CC compositions and vaccines for mycobacteria other than of the  
 CC M. tuberculosis complex, e.g. M. avium which infects poultry and  
 CC occasionally humans, M. leprae; they are especially useful when they do  
 CC not react with lymphoid cells previously primed with M. tuberculosis  
 CC complex mycobacteria, and so do not give rise to a diagnostic reaction  
 CC in individuals infected with these bacteria. The polypeptides may also  
 CC be used in in vitro diagnostic tests e.g. stimulation of interferon-gamma  
 CC release from lymphocytes. The polypeptide has similar or higher  
 CC bioreactivity as currently used tuberculin reagent in the standard  
 CC delayed type hypersensitivity (DTH) skin test for tuberculosis, but may  
 CC have greater specificity, being better able to discriminate between  
 CC lymphoid cell primed from tuberculosis and from previous vaccination.  
 CC The present sequence represents M. tuberculosis ESAT-6 used in the  
 CC exemplification of the present invention.

XX Sequence 95 AA:  
 Query Match 100.0%; Score 77; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 1e-05; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservation 0; MisMatches 0; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 ||||||| | | | | | | | |  
 6 WNFAGIEAAASATQG 20  
 Db  
 RESULT 14  
 AAY29788  
 ID AAY29788 standard; Protein: 95 AA.  
 AC AAY29788;  
 XX DT 08-NOV-1999 (first entry)  
 DE Mycobacterial tuberculosis ESAT-6 protein.  
 KW Mycobacterial; ESAT-6; 6 kDa antigen; tuberculosis; infection;  
 KW immunological response; diagnosis; vaccine.  
 XX OS Mycobacterium tuberculosis.  
 XX US5955077-A.  
 PN 21-SEP-1999.  
 PR 05-JUN-1995; 95US-0465640.  
 PR 05-JUN-1995; 95US-0465640.  
 PR 20-SEP-1993; 93US-0123182.  
 PR 01-JUL-1994; 94WO-DK00273.  
 PA (STAT-) STATENS SERUMINSTITUT.  
 XX PT (STAT-) STATENS SERUMINSTITUT.  
 XX PI Andersen AB, Andersen P, Haslov K, Sorensen AL;  
 XX DR N-PADB; AAZ08877.  
 XX WPI; 1999-539545/45.  
 XX Polypeptide secreted from Mycobacterium is useful as a vaccine  
 XX against tuberculosis  
 XX claim 24; FIG 10c; 39pp; English.  
 XX  
 CC The present invention describes a purified or non-naturally occurring  
 CC polypeptide (I) released from a metabolising mycobacteria comprising an  
 CC ESAT6, also called the 6 kDa antigen. The present sequence represents  
 CC Mycobacterium tuberculosis ESAT-6 protein. Also described is a purified  
 CC or non naturally occurring Polypeptide (II) with the ability to elicit  
 CC a delayed type hypersensitivity reaction which comprises a T cell  
 CC

CC epitope of (I). (II) can be used with a carrier or vehicle in a  
 CC composition for diagnosing tuberculosis caused by mycobacteria belonging  
 CC to the tuberculosis complex, i.e. Mycobacterium tuberculosis, M. bovis  
 CC and M. africanum. The composition can be used to detect microbial  
 CC antibodies or components of mycobacteria in samples or in animals  
 CC through the use of immunoassays. (II) can be used as a vaccine for  
 CC immunising an animal, including humans against tuberculosis caused by  
 CC mycobacteria of the tuberculosis-complex. (II) induce a release of  
 CC IFN-gamma from reactivated T-lymphocytes evoking a protective immune  
 CC response. Vaccine containing (II) has the same protective potency as  
 CC the live BCG vaccine against tuberculosis.

XX Sequence 95 AA:  
 Query Match 100.0%; Score 77; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 1e-05; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservation 0; MisMatches 0; Indels 0; Gaps 0;  
 QY 1 WNFAGIEAAASATQG 15  
 ||||||| | | | | | | | |  
 6 WNFAGIEAAASATQG 20  
 Db  
 RESULT 15  
 AAB35219  
 ID AAB35219 standard; Protein: 95 AA.  
 AC AAB35219;  
 XX DT 24-APR-2001 (first entry)  
 XX DE Mycobacterium Rv3875 protein.  
 XX KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036C;  
 KW Rv1037C; Rv2346C; Rv2653C; Rv3020C; Rv3444C;  
 KW Rv3445C; Rv3890C; Rv3904C; Rv3905C;  
 XX OS Mycobacterium tuberculosis.  
 XX PD 18-JAN-2001.  
 XX PF WO20010151-A2.  
 XX PR 13-JUL-2000; 99DK-0001020.  
 PR 15-JUL-1999; 99US-0144011.  
 XX PA (STAT-) STATENS SERUM INST.  
 XX PT Andersen P, Skjot R;  
 XX DR WPI; 2001-091923/10.  
 XX PT New polypeptide encoded by a member of the esat-6-gene family for  
 XX PT immunizing against and diagnosis of tuberculosis -  
 XX PS Example 2; Page 65; 80pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for  
 CC members of the esat-6 gene family from Mycobacterium tuberculosis. These  
 CC proteins include Rv0287, Rv1036C, Rv1037C, Rv2346C, Rv348C, Rv2653C,  
 CC Rv2654C, Rv3020C, Rv3445C, Rv3890C, Rv3891C, Rv3904C and  
 CC Rv3905C. These can be used to produce vaccines against, and in the  
 CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of  
 CC the proteins of the invention.

XX Sequence 95 AA;  
 Query Match 100.0%; Score 77; DB 22; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Jul 3 14:50:02 2003

us-09-830-839-6.rag

Page 8

Oy 1 WNFAGIERAASA1QG 15  
Db 6 WNFAGIERAASA1QG 20

Search completed: July 3, 2003, 14:05:32  
Job time : 31 secs



; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag985B  
**US-09-805-427A-4**

Query Match, Best Local Similarity 100.0%; Pred. No. 1.4e-05; Score 79; DB 9; Length 403;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
**Oy** 1 MTBQWNFAGIEAAA 15  
 Db 22 MTBQWNFAGIEAAA 36

RESULT 3

**US-09-791-171-173**  
 Sequence 173, Application US/09791171  
 Patent No. US20020094336A1

GENERAL INFORMATION:  
 APPLICANT: ANDERSEN, Peter

APPLICANT: NIELSEN, Rikke

APPLICANT: OETTINGER, Thomas

APPLICANT: RASMUSSEN, Peter Birk

APPLICANT: ROSENKRANDS, Ida

APPLICANT: WELDINGH, Karin

APPLICANT: FLORIO, Walter

TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS

FILE REFERENCE: 670001-2002.1  
 CURRENT APPLICATION NUMBER: US/09/791,171

CURRENT FILING DATE: 2001-02-20  
 PRIOR APPLICATION NUMBER: 09/050,739

PRIOR FILING DATE: 1998-03-30  
 PRIOR APPLICATION NUMBER: 03/66/97

PRIOR FILING DATE: 1997-04-02  
 PRIOR APPLICATION NUMBER: 12/77/97

PRIOR FILING DATE: 1997-11-10  
 PRIOR APPLICATION NUMBER: 60/044,624

PRIOR FILING DATE: 1997-04-18  
 PRIOR APPLICATION NUMBER: 60/070,488

PRIOR FILING DATE: 1998-01-05  
 NUMBER OF SEQ ID NOS: 173

SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 173  
 LENGTH: 403

TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis

**US-09-791-171-173**

Query Match, Best Local Similarity 100.0%; Pred. No. 1.4e-05; Score 79; DB 9; Length 403;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Oy**

1 MTBQWNFAGIEAAA 15  
 Db 22 MTBQWNFAGIEAAA 36

RESULT 4

**US-09-805-427A-3**  
 Sequence 3, Application US/09805427A  
 Patent No. US2002016867A1

GENERAL INFORMATION:  
 APPLICANT: Statens Serum Institut

TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens

FILE REFERENCE: 670001-002.5  
 CURRENT FILING DATE: 2001-03-13  
 NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 3  
 LENGTH: 404

TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis

**US-09-791-171-172**

Query Match, Best Local Similarity 100.0%; Pred. No. 1.4e-05; Score 79; DB 9; Length 403;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Oy**

1 MTBQWNFAGIEAAA 15  
 Db 310 MTBQWNFAGIEAAA 324

RESULT 5

**US-09-791-171-172**  
 Sequence 172, Application US/09791171  
 Patent No. US20020094336A1

GENERAL INFORMATION:  
 APPLICANT: ANDERSEN, Peter

APPLICANT: NIELSEN, Rikke

APPLICANT: OETTINGER, Thomas

APPLICANT: RASMUSSEN, Peter Birk

APPLICANT: ROSENKRANDS, Ida

APPLICANT: WELDINGH, Karin

APPLICANT: FLORIO, Walter

TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS

FILE REFERENCE: 670001-2002.1  
 CURRENT APPLICATION NUMBER: US/09/791,171

CURRENT FILING DATE: 2001-02-20  
 PRIOR APPLICATION NUMBER: 09/050,739

PRIOR FILING DATE: 1998-03-30  
 PRIOR APPLICATION NUMBER: 03/66/97

PRIOR FILING DATE: 1997-04-02  
 PRIOR APPLICATION NUMBER: 12/77/97

PRIOR FILING DATE: 1997-11-10  
 PRIOR APPLICATION NUMBER: 60/044,624

PRIOR FILING DATE: 1997-04-18  
 PRIOR APPLICATION NUMBER: 60/070,488

PRIOR FILING DATE: 1998-01-05  
 NUMBER OF SEQ ID NOS: 173

SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 172  
 LENGTH: 404

TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis

**US-09-791-171-172**

Query Match, Best Local Similarity 100.0%; Pred. No. 1.4e-05; Score 79; DB 10; Length 404;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Oy**

1 MTBQWNFAGIEAAA 15  
 Db 310 MTBQWNFAGIEAAA 324

RESULT 6

**US-10-044-703-60**  
 Sequence 60, Application US/10044703  
 Publication No. US20020192233A1

GENERAL INFORMATION:  
 APPLICANT: De groot, Anne S

TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters  
 FILE REFERENCE: 17999-004 US  
 CURRENT APPLICATION NUMBER: US/10/044,703  
 PRIOR APPLICATION NUMBER: 60/190,834  
 PRIOR FILING DATE: 2002-05-20  
 PRIOR APPLICATION NUMBER: 60/190,834  
 PRIOR FILING DATE: 2000-03-20  
 NUMBER OF SEQ ID NOS: 81  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 60

RESULT 7  
US-09-813-333-60  
; Sequence 60, Application US/09813333  
; Patent No. US20020119160A1

GENERAL INFORMATION:  
APPLICANT: Degroot, Anne S  
TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters  
FILE REFERENCE: 17999-004 US  
CURRENT APPLICATION NUMBER: US/09/813, 333

CURRENT APPLICATION NUMBER: US/09/813, 333  
CURRENT FILING DATE: 2001-03-20  
PRIORITY FILING DATE: 2000-03-20  
PRIORITY APPLICATION NUMBER: 60/190, 834  
PRIOR FILING DATE: 2000-03-20  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 60  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis

US-09-813-333-60

Query Match 74.7%; Score 59; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.0017; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 QWNFAGIEAAA 15  
Db 1 QWNFAGIEAAA 11

RESULT 8  
US-10-044-703-59  
; Sequence 59, Application US/10044703  
; Publication No. US20020192233A1

GENERAL INFORMATION:  
APPLICANT: Degroot, Anne S  
TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters  
FILE REFERENCE: 17999-004 US  
CURRENT APPLICATION NUMBER: US/10/044, 703

CURRENT APPLICATION NUMBER: US/10/044, 703  
CURRENT FILING DATE: 2002-05-20  
PRIORITY FILING DATE: 2000-03-20  
PRIORITY APPLICATION NUMBER: 60/190, 834  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 61  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis

US-10-044-703-59

Query Match 74.7%; Score 59; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.0017; Mismatches 0; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 QWNFAGIEAAA 15  
Db 1 QWNFAGIEAAA 11

RESULT 9  
US-09-813-333-61  
; Sequence 61, Application US/09813333  
; Patent No. US20020119160A1

GENERAL INFORMATION:  
APPLICANT: Degroot, Anne S  
TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters  
FILE REFERENCE: 17999-004 US  
CURRENT APPLICATION NUMBER: US/09/813, 333

CURRENT APPLICATION NUMBER: US/09/813, 333  
CURRENT FILING DATE: 2001-03-20  
PRIORITY FILING DATE: 2000-03-20  
PRIORITY APPLICATION NUMBER: 60/190, 834  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 61  
LENGTH: 20

RESULT 9  
US-09-813-333-61  
; Sequence 61, Application US/09813333  
; Patent No. US20020119160A1

GENERAL INFORMATION:  
APPLICANT: Degroot, Anne S  
TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters  
FILE REFERENCE: 17999-004 US  
CURRENT APPLICATION NUMBER: US/09/813, 333

CURRENT APPLICATION NUMBER: US/09/813, 333  
CURRENT FILING DATE: 2001-03-20  
PRIORITY FILING DATE: 2000-03-20  
PRIORITY APPLICATION NUMBER: 60/190, 834  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 61  
LENGTH: 20

RESULT 9

TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
; US-09-813-333-61

Query Match Score 54.4%; Score 43; DB 10; Length 20;  
; Best Local Similarity 100.0%; Pred. No. 0.84; 0; Mismatches 0; Indels 0; Gaps 0;  
; Matches 9; Conservative 0;

Qy 7 NFAGIEAAA 15  
Db 1 NFAGIEAAA 9

RESULT 12  
US-09-738-626-4782  
; Sequence 4782, Application US/09738626  
; Publication No. US20020197605A1

GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patentin ver. 3.0  
; SEQ ID NO 4782  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
; US-09-738-626-4782

RESULT 14  
US-09-738-626-6880  
; Sequence 6880, Application US/09738626  
; Publication No. US20020197605A1

GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patentin ver. 3.0  
; SEQ ID NO 6880  
; LENGTH: 301  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
; US-09-738-626-6880

Query Match Score 49.4%; Score 39; DB 10; Length 916;  
; Best Local Similarity 54.5%; Pred. No. 1.9e+02; 2; Mismatches 2; Indels 0; Gaps 0;  
; Matches 6; Conservative 3;

Qy 4 QDWNFAGIEAA 14  
Db 699 EQMDIEGLEAA 709

RESULT 13  
US-09-815-242-12006  
; Sequence 12005, Application US/09815242  
; Patent No. US20020051569A1

GENERAL INFORMATION:  
; APPLICANT: NTEQQWNEA 9  
; LENGTH: 58

Query Match Score 50.6%; Score 40; DB 9; Length 295;  
; Best Local Similarity 55.6%; Pred. No. 40; 0; Mismatches 0; Indels 0; Gaps 0;  
; Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTEQQWNEA 9  
Db 50 LTDEEWWYA 58

RESULT 13  
US-09-815-242-12006  
; Sequence 12005, Application US/09815242  
; Patent No. US20020051569A1

GENERAL INFORMATION:  
; APPLICANT: Hselseck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Irwick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.01A  
; CURRENT APPLICATION NUMBER: US/09/815, 242  
; CURRENT FILING DATE: 2001-03-21

RESULT 15

US-09-847-081B-4

; Sequence 4 , Application US/09847081B

; Patent No. US20020128464A1

; GENERAL INFORMATION:

; APPLICANT: BAYER AG

; TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase

; FILE REFERENCE: La A 34 326

; CURRENT APPLICATION NUMBER: US/09/847,081B

; CURRENT FILING DATE: 2001-05-02

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 410

; TYPE: PRT

; ORGANISM: Nicotiana tabacum

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 135, 139

; OTHER INFORMATION: xaa is unknown or other

; US-09-847-081B-4

Query Match 48.1%; Score 38; DB 10; Length 410;

Best Local Similarity 41.7%; Pred. No. 1 2e-02;

Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 EQWNIFAGIBAA 14

Db 5 0 KQRWNIFGGSVRSAA 61

Search completed: July 3, 2003, 14:08:27  
Job time : 18.5 secs



Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:03:09 ; (without alignments) search time 24 Seconds

Scoring table: BLOSUM62 Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries

Database : SPREMBL\_21:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mmc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query	Match	Length	DB	ID	Description
1	43	54.4	217	16	O9KMYB	Q9KMY8 vibrio chol	
2	43	54.4	247	16	O8X3FB	Q8X3FB escherichia	
3	42	53.2	369	16	O9Z839	Q9Z839 chlamydia p	
4	42	53.2	723	16	O9A279	Q9a279 Caulobacter	
5	41	51.9	182	2	P95622	P95622 rhizobium 1	
6	41	51.9	440	2	Q9RAD1	Q9fad1 moritella J	
7	41	51.9	443	2	O9AMK7	O9amk7 vibrio vuln	
8	41	51.9	445	16	O9KV76	O9kv76 vibrio chol	
9	41	51.9	471	17	O28786	O28786 archaeoglob	
10	41	51.9	934	16	O8XVJ6	O8xvj6 ralstonia s	
11	40	50.6	16	16	O9TKC4	O9tkc4 vibrio chol	
12	39	49.4	81	16	O981F7	O981f7 rhizobium 1	
13	39	49.4	219	16	O8YVX4	O8yvx4 anaerana sp	
14	39	49.4	282	2	O8RQL0	O8rql0 pasteurella	
15	39	49.4	325	5	O966M0	O966m0 caenorhabdi	
16	39	49.4	328	16	051337	051337 borrelia bu	

**ALIGNMENTS**

RESULT 1

O9KMY8 PRELIMINARY; PRT; 217 AA.

ID O9KMY8; Q9KMY8; DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE FRNE Protein.

GN VCA0178.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE 01;

RX MEDLINE=200406833; PUBMED=10952001;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.B., Read T.D., Rettelein H., Richardson D.,

RA Emiolaeva M.D., Vamathevan J., Bass S.S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S. L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.,

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.",

RL Nature 406:477-483 (2000);

DR EMBL: AE004358; AAF6091.1; -.

DR TIGR: VCA0178; -.

DR Intertaxa: IPR004287; HCCA\_isomerase.

DR Pfam: PF03046; HCCA\_isomerase; 1.

KW Complete proteome.

SO SEQUENCE 217 AA; 24897 MW; 79B473E32313B5B8 CRC64;

Query Match 54.4%; Score 43; DB 16; Length 217;

Best Local Similarity 75.0%; Pred. No. 14;

Matches 9; Conservative 75.0%; Mismatches 0; Indels 3; Gaps 0;

DB	170 TEQOWLQAGIHA 181	RL Nucleic Acids Res. 28:1397-1406(2000). [3]
RN		SEQUENCE FROM N.A.
RC		STRAIN-J138;
RX		MEDLINE=2030349; PubMed=10871362;
ID	Q9X3F8	Shirai M., Hirakawa K., Tabuchi M., Kishi F., Ouchi K., Shibata T., Ishii K., Hattori M., Kuwara S., Nakazawa T.; Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA;"
AC	Q9X3F8	RT Nucleic Acids Res. 28:2311-2314(2000);
DT	01-MAR-2002 (TREMBREL, 20, Last sequence update)	DR EMBL; AE001636; RAD18653; 1; -;
DT	01-JUN-2002 (TREMBREL, 21, Last annotation update)	DR EMBL; AE002184; AAF38105; 1; -;
DE	Hypothetical protein Ecs1576.	DR EMBL; CP02546; BAA98719; 1; -;
GN	Ecs1576	DR TIGR; CP0240; -;
OS	Escherichia coli O157:H7.	DR InterPro; IPRO05244; Cons_hypoth423.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	DR TIGRFAMS; TIGR00423; TIGR00423; 1.
OX	NCBI_TaxID:83334;	KW Complete proteome.
RN	[1]	SQ SEQUENCE FROM N.A.: 247 AA; 723 MW; 05AC33C15479B192 CRC64;
RC		STRAIN=O157:H7 / RIMD 050952;
RX		MEDLINE=21156231; PubMed=11258796;
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,	RA Shiba T., Ishii K., Hattori M., Kuwara S., Nakazawa T.; Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA;"
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,	RT Nucleic Acids Res. 28:2311-2314(2000);
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,	DR EMBL; AE001636; RAD18653; 1; -;
RA	Kunbara S., Shiga T., Hattori M., Shinagawa H.;	DR EMBL; AE002184; AAF38105; 1; -;
RT	"Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";	DR EMBL; CP02546; BAA98719; 1; -;
RL	DNA Res. 8:1-22(2001).	DR TIGR; CP0240; -;
DR	EMBL; AP002555; BAB3499; 1; -;	DR InterPro; IPRO05244; Cons_hypoth423.
DR	InterPro; IPRO00508; SigPbase.	DR TIGRFAMS; TIGR00423; TIGR00423; 1.
DR	PROSITE; PS00761; SPASE_I_3; UNKNOWN_1.	KW Complete proteome.
KW	Hypothetical protein.	SQ SEQUENCE FROM N.A.: 369 AA; 41556 MW; 05AC33C15479B192 CRC64;
SO		STRAIN=O157:H7 / RIMD 050952;
Query Match	54.4%	Score 43; DB 16; Length 247;
Best Local Similarity	53.8%	Score 43; DB 16; Length 247;
Matches	7; Conservative	Score 43; DB 16; Length 247;
Matches	7; Conservative	Pred. No. 16; Pred. No. 16; Gaps 0;
QY	3 EDQWNFGIEAAA 15	QY 6 WNTQFGEAAA 14
OS	:  :  :  :  :	OS 305 WNTQFGEAAA 313
Db	22 QDNWGFGSGKKA 34	Db
RESULT 3		RESULT 4
Q9Z839	PRELIMINARY; PRT; 369 AA.	Q9A279 PRELIMINARY; PRT; 723 AA.
ID	Q9Z839	ID Q9A279
AC	Q9Z839; 01-MAY-1999 (TREMBREL, 10, Created)	AC Q9A279;
DT	01-MAY-1999 (TREMBREL, 10, Last sequence update)	DT 01-JUN-2001 (TREMBREL, 17, Last sequence update)
DR	01-MAY-2002 (TREMBREL, 21, Last annotation update)	DR 01-JUN-2001 (TREMBREL, 20, Last annotation update)
DE	FE-S oxidoreductase.	DE Prolyl oligopeptidase family protein.
GN	Cpn0513 OR CP0513 OR CP0240	GN CC3687.
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).	OS Caulobacter crescentus
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.	OC Caulobacter
OC	NCBI_TaxID:83558;	OC Caulobacter
OX		OX NCBI_TaxID:155892;
RN	[1]	RN [1]
RP	SEQUENCE FROM N.A.	RP SEQUENCE FROM N.A.
RC		RC STRAIN=ATCC 19089 / CB15;
RX		RX MEDLINE=21173698; PubMed=11259647;
RA	Nierman W.C., Feilberg T.W., Laub M.T., Paulsen I.T., Nelson K.E.,	RA Elsen J., Heidebrink J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Hart D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Emalava M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M., Ratliff J., "Complete genome sequence of Caulobacter crescentus.," Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	DR EMBL; AE006026; AAK25649; 1; -;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	DR TIGR; CC3687; -;
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,	DR InterPro; IPRO01375; Peptidase_S9.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO04106; Peptidase_S9_N.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO02470; Prolico_prase.
RT	Nat. Genet. 21:385-389(1999).	DR InterPro; IPRO00379; Ser_estr_sit.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR Peptidase_S9; Peptidase_S9_N; 1.
RC		DR PRINTS; PR00862; PROLIGO_PTASE.
RX		KW Complete proteome.
RA	Medline=20150255; Pubmed=10684935;	SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=20150255; Pubmed=10684935;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999).	DR Peptidase_S9; Peptidase_S9_N; 1.
RN	[2]	DR Peptidase_S9; Peptidase_S9_N; 1.
RP	SEQUENCE FROM N.A.	DR PRINTS; PR00862; PROLIGO_PTASE.
RC		KW Complete proteome.
RX		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RA	Medline=99206606; Pubmed=10192388;	Query Match 53.2%; Score 42; DB 16; Length 723;
RA	Read T.D., Brunham R.C., Shen C.R., Heidelberg J.F.,	Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Gwin M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Eisen J., Fraser C.M.,	DR Peptidase_S9; Peptidase_S9_N; 1.
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	DR PRINTS; PR00862; PROLIGO_PTASE.
RT	pneumoniae RR39.,"	KW Complete proteome.
RN		SQ SEQUENCE FROM N.A.: 723 AA; 78547 MW; 05FB0515D74819AB CRC64;
RC		Query Match 53.2%; Score 42; DB 16; Length 723;
RX		Best Local Similarity 50.0%; Pred. No. 72; Indels 0; Gaps 0;
RA	Medline=99206606; Pubmed=10192388;	Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
RA	Kalmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	DR Peptidase_S9; Peptidase_S9_N.
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,	DR InterPro; IPRO02470; Prolico_prase.
RT	Comparative genomics of Chlamydia pneumoniae and C. trachomatis.,"	DR InterPro; IPRO00379; Ser_estr_sit.
RT	Nat. Genet. 21:385-389(1999	

**RESULT 5**

ID P95622	PRELIMINARY;	PRT; 182 AA.
AC P95622;		
DT 01-MAY-1997 (TREMBlrel. 03, Created)		
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)		
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)		
DE Cytochrome c1 (Fragment).		
OS Rhizobium leguminosarum.		
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC Rhizobiaceae; Rhizobium.		
OX NCBI_TaxID=384;		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE=97158229; PUBMED=9004501;		
RA Wu G., Delgado M.J., Vargas C., Davies A.E., Poole R.K., Downie J.A.;		
RT "The cytochrome bcl complex but not CycCm is necessary for symbiotic nitrogen fixation by Rhizobium leguminosarum.";		
RL Microbiology 142:3381-3388 (1996).		
DR EMBL: X88018; CAA66511.; -		
DR InterPro: IPR00345; CYTC_heme_bind.		
DR InterPro: IPR002326; CYT_Cl.		
DR Pfam: PF02167; Cytochrome_C1.; 1.		
DR PRINTS: PR00603; CYTOCHROMECL.		
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.		
FT NON_TER 182		
SQ SEQUENCE 182 AA; 19787 MW; BC1C7255859D44B8 CRC64;		
Query Match 51.9%; Score 41; DB 2; Length 182;		
Best Local Similarity 66.7%; Pred. No. 26; Mismatches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		

**Query 2 TEQQWNFAAG 10**

Db 51 TEEWNTFAAG 59		
---------------------	--	--

**RESULT 6**

ID Q9FAD1	PRELIMINARY;	PRT; 440 AA.
ID Q9FAD1		
AC Q9FAD1;		
DT 01-MAR-2001 (TREMBlrel. 16, Created)		
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)		
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE CadB.		
GN CADB.		
OS Moritella japonica.		
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;		
OC Moritella.		
OX NCBI_TaxID=89067;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=DS1;		
RA Li L., Fujii S., Kato C., Koid H.;		
RT "Expression of Lysin Decarboxylase Gene Is Regulated by Pressure in a Deep-Sea Piezophilic Bacterium, Moritella japonica.";		
RT Submitted (APR 2000) to the EMBL/GenBank/DDBJ databases.		
DR EMBL: AB041735; BAB16717.; -		
DR InterPro: IPR002293; AA_rel_permeasel.		
DR InterPro: IPR004841; Permease.		
DR Pfam: PF00324; aa_permeases; 1.		
SQ SEQUENCE 440 AA; 45577 MW; DBDE4CADF8958CD9 CRC64;		
Query Match 51.9%; Score 41; DB 2; Length 440;		
Best Local Similarity 60.0%; Pred. No. 65; Mismatches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		

**Query 6 WNFAGTEAAA 15**

Db 199 WNFVGVESAA 208		
-----------------------	--	--

**RESULT 7**

ID Q9AMK7	PRELIMINARY;	PRT; 443 AA.
ID Q9AMK7		
AC Q9AMK7;		
DT 01-JUN-2001 (TREMBlrel. 17, Created)		
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)		
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)		
DE Cadaverine/Lysine antiporter.		
OS Vibrio vulnificus.		
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.		
OX NCBI_TaxID=672;		
RN [1]		
RP SEQUENCE FROM N.A.		
RA Rhee J., Lee J., Lee S., Chung S., Rhee J., Choi S., Ryu P.;		
RT "Vibrio vulnificus cadBA genes.";		
RT Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.		
DR EMBL: AF334470; AAC01663.1.; -		
DR InterPro: IPR002293; AA_rel_permeasel.		
DR InterPro: IPR004841; Permease.		
DR Pfam: PF00324; aa_permeases; 1.		
SQ SEQUENCE 443 AA; 46752 MW; B88FD7E188036751 CRC64;		
Query Match 51.9%; Score 41; DB 2; Length 443;		
Best Local Similarity 60.0%; Pred. No. 65; Mismatches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		

**Query 6 WNFAGTEAAA 15**

Db 199 WNFVGVESAA 208		
-----------------------	--	--

**RESULT 8**

ID Q9KV76	PRELIMINARY;	PRT; 445 AA.
ID Q9KV76		
AC Q9KV76;		
DT 01-OCT-2000 (TREMBlrel. 15, Created)		
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DE Cadaverine/lysine antiporter CadB, putative.		
GN VC0280.		
OS Vibrio cholerae.		
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.		
OX NCBI_TaxID=66;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=EL TOR NI6961 / SEROTYPE O1;		
RA MEDLINE=20406833; PubMed=10952301;		
RA Heidelberg J.F., Elsen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,		
RA Dossen R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,		
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,		
RA Brimholmaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I., Sellers P.,		
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,		
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,		
RA Fraser C.M.,		
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";		
RT Nature 405:477-483 (2000).		
DR TIGR: VC0280; -		
DR InterPro: IPR002293; AA_rel_permeasel.		
DR InterPro: IPR004841; Permease.		
DR Pfam: PF00324; aa_permeases; 1.		
SQ SEQUENCE 445 AA; 46907 MW; 3A64F21B9680694C CRC64;		
Query Match 51.9%; Score 41; DB 16; Length 445;		
Best Local Similarity 60.0%; Pred. No. 65; Mismatches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;		

**Query 6 WNFAGTEAAA 15**

Db 199 WNFVGVESAA 208		
-----------------------	--	--

Db 199 WSFGVGSAA 208

RESULT 9

ID 028786 PRELIMINARY; PRT; 471 AA.

AC 028786; 1998 (TREMBrel. 05, Created)

DT 01-JAN-1998 (TREMBrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBrel. 20, Last annotation update)

DE Hypothetical protein AF1486.

GN AF1486.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OX NCBI\_TaxID=2234;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Richardson D.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Fleischmann R.D., Kerlavage A.R., Graham D.E., Kyrpides N.C., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

RT "the complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997);

DR EMBL; AE001000; AAC88769.1; -.

DR TIGR; AF1486; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 471 AA; 5416 MW; 892A03A47FBDA71 CRC64;

Query Match 51.9%; Score 41; DB 17; Length 471;

Best Local Similarity 61.5%; Pred. No. 69;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 EDQWNFACTEAAA 15

DB 45 E0WDFFEGKEADA 57

RESULT 10

RN 08XVJ6 . PRELIMINARY; PRT; 934 AA.

AC 08XVJ6; 2002 (TREMBrel. 20, Created)

DT 01-MAR-2002 (TREMBrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBrel. 21, Last annotation update)

DE Probable preprotein translocase SECA subunit.

GN SECA OR RSC2834 OR RS00271.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.

OX NCBI\_TaxID=305;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GM1000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Aillet M., Billault A., Brottier P., Camus J.C., Cattolico L., Chandler M., Choise N., Claudel-Renard S., Cunac S., Demange N., Gaspin C., Lavia M., Moisan A., Robert C., Saurin W., Schleiff T., Siguiier P., Thebaud P., Whalen M., Wincker P., Levy M., Weissenbach J., Boucher C.A.;

"Genome sequence of the plant pathogen Ralstonia solanacearum."; Nature 415:497-502(2002).

DR EMBL; AL646072; CAD16541.1; -.

DR Interpro; IPR00185; SecA.

DR Interpro; IPR004027; SEC-C\_motif.

DR Pfam; PF02810; SEC-C; 1.

DR PRINTS; PR01043; SEC-A\_protein; 1.

DR TIGRFAMS; TIGR00963; secA; 1.

DR PROSITE; PS01312; SEC4; 1.

KW Complete proteome.

SQ SEQUENCE 934 AA; 105241 MW; F9C3FA601EA62095 CRC64;

Query Match Best Local Similarity 51.9%; Score 41; DB 16; Length 934; Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 EDQWNFACTEAAA 12

Db 708 EEQDNDAGIE 717

RESULT 11

RP Q9TC4 PRELIMINARY; PRT; 38 AA.

ID Q9TC4; QRTIC4

AC 01-OCT-2000 (TREMBrel. 15, Created)

DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)

DE Hypothetical protein VC0978.

GN VC0978.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR NI6961 / SEROTYPE 01;

RX MEDLINE=2046833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Emrhaeva M.D., Yamamoto J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;

RA RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";

RL Nature 406:477-483(2000);

DR EMBL; AE004179; AAF94140.1; -.

DR TIGR; VC0978; -.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 38 AA; 4427 MW; 30774987004F910 CRC64;

Query Match Best Local Similarity 50.6%; Score 40; DB 16; Length 38;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTESQWNFACTEAAA 14

DB 18 MTLNQWNWIGVGVA 31

RESULT 12

RN 098IF7 PRELIMINARY; PRT; 81 AA.

AC 098IF7; 098IF7;

DT 01-OCT-2001 (TREMBrel. 18, Created)

DT 01-OCT-2001 (TREMBrel. 18, Last sequence update)

DT 01-OCT-2002 (TREMBrel. 20, Last annotation update)

DE Hypothetical protein msr2423.

GN MSR2423.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OX NCBI\_TaxID=381;

Mesorhizobium.

RN [1] SQUENCE FROM N.A.  
 RP  
 RC STRAIN=MAF303059;  
 RX MEDLINE=21082939; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Kishida T., Kiyokawa C., Kohara M., Kawashima K., Kimura T.,  
 RA Matsunoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RP "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AP003999; BAB49559.1.  
 DR InterPro; IPR002145; CopG\_RTH\_4.  
 DR Pfam; PF01402; HTH\_4; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 81 AA; 9035 MW; 7652A22945D0BFC7 CRC64;  
 Query Match 49.4%; Score 39; DB 16; Length 81;  
 Best Local Similarity 53.8%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 MTBQWQFAGIEA 13  
 Db 41 VTRQEMQLAIEA 53

RESULT 13  
 Q8YVX4 PRELIMINARY; PRT; 219 AA.  
 ID Q8YVX4  
 AC 08YVX4; Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein Alr1845.  
 GN ALR1845.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OC NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595283; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsunoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RP "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213 (2001).  
 DR EMBL; AP003587; BAB3544.1; -.  
 DR InterPro; IPR004395; Cons\_hypoth91.  
 DR InterPro; IPR003358; Methyltransf\_4.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR Pfam; PF02390; Methytransferf\_4; 1.  
 DR TIGRFMS; TIGR00091; Cons\_hypoth91; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 219 AA; 24847 MW; 6B25C82EA25C83C CRC64;  
 Query Match 49.4%; Score 39; DB 16; Length 219;  
 Best Local Similarity 60.0%; Pred. No. 71; Mismatches 6; Conservative 6; Indels 0; Gaps 0;  
 Matches 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 14  
 Q8RQLO PRELIMINARY; PRT; 282 AA.  
 ID Q8RQLO  
 AC 08RQLO;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)

DR 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 RC Beta lactamase.  
 RA STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RA "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* Sequencing Consortium.;"  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Wansley P.;  
 RT "The sequence of *C. elegans* cosmid C55B7.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005625; AAK6871.1; -.  
 SQ SEQUENCE 325 AA; 37141 MW; 55EAM41865C645C0 CRC64;  
 Query Match 49.4%; Score 39; DB 5; Length 325;  
 Best Local Similarity 46.2%; Pred. No. 1.1e+02; Mismatches 6; Conservative 6; Indels 4; Gaps 0;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EQQWNFAGIEAA 15  
 Db 299 EEQWNLSATEATS 311

Search completed: July 3, 2003, 14:06:47  
 Job time : 27 sec

Thu Jul 3 14:50:02 2003

us-09-830-839-1.rspt

OM protein - protein search, using sw model

Run on: July 3, 2003, 14:00:24 ; search time 30 Seconds  
 (without alignments)  
 66.625 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79 MTEQQNNFAGIEAAA 15

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDS2/gcdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDS2/gcdata/geneseq/geneseq-emb1/AA2001.DAT:\*

22: /SIDS2/gcdata/geneseq/geneseq-emb1/AA2002.DAT:\*

23: /SIDS2/gcdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	DB ID
1	79	100.0	15 21 AAY94590
2	79	100.0	15 23 AAY017432
3	79	100.0	17 21 AAY80581
4	79	100.0	19 18 AAW5548
5	79	100.0	18 AAW2466
6	79	100.0	51 18 AAW3339
7	79	100.0	51 19 AAW81701
8	79	100.0	51 19 AAW64334
9	79	100.0	51 20 AAY39131
10	100.0	51	20 AAY38988

RESULT 1

ID	AYY94590 standard; peptide; 15 AA.
XX	AC AAY94590;
XX	DT 28-NOV-2000 (first entry)
XX	DE Mycobacterium tuberculosis ESAT-6 peptide E51.
XX	KW Tuberculosis; infection diagnosis; ESAT-6.
XX	OS Mycobacterium tuberculosis.
PN	WO2000026248-A2.
XX	PD 11-MAY-2000.
XX	PF 03-NOV-1999; 99WO-GB03635.
PR	04-NOV-1998; 98GB-0024213.
PR	04-NOV-1998; 98US-0107004.
PA	(ISIS-) ISIS INNOVATION LTD.
XX	PT Lalvani A, Pathan AA;
XX	DR WPI; 2000-365579/31.
PT	Novel method of diagnosing infection, or exposure of a host, to a mycobacterium comprising contacting T cells from the host with ESAT-6 derived peptides
XX	M. tuberculosis ES

PS Claim 1; Page 3; 33pp; English.

CC Mycobacterium tuberculosis useful in the method of the invention.

CC The present sequence is the peptide ESI derived from the *Mycobacterium*

CC *tuberculosis* ESAT-6 gene. This sequence is one of eleven peptides

CC derived from the ESAT-6 gene (see AAV94590 to AAV9600). The peptides

CC are recognised, to varying degrees, by the T cells of patients with

CC tuberculosis. When the peptides are combined together as a panel they

CC provide a high specificity and sensitivity diagnostic test for

CC *M. tuberculosis* infection. This test has the advantage that it does not

CC give false positives for patients who have been vaccinated with BCG.

XX Sequence 15 AA;

SQ Query Match 100.0%; Score 79; DB 21; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEQQWNFAGIEAA 15

Db 1 MTEQQWNFAGIEAA 15

RESULT 2

AAO17432

AAO17432 standard; peptide: 15 AA.

AC XX AAO17432;

DT 11-JUL-2002 (first entry)

XX DE Early secreted antigenic target 6 T cell epitope #1.

XX KW Tuberculosis; TB; vaccination; vaccine; CD4+ T cell immune response;

XX KW poxvirus; vector; HIV; malaria; *Helicobacter pylori*; influenza;

XX KW hepatitis; viral infection; leprosy; protozoan parasite; cancer;

XX KW tubercostatic; anti-HIV; protozoacide; antibacterial; virucide;

XX KW hepatotropic; antiinflammatory; antileprotic; cytostatic; epitope.

OS *Mycobacterium tuberculosis*.

OS XX

PN WO200224224-A2.

PN XX

PD 28-MAR-2002.

PD XX

PF 13-SEP-2001; 2001WO-GB04116.

PF XX

PR 21-SEP-2000; 2000GB-0023203.

PR XX

PA (OXO-) OXON PHARMACEINES LTD.

PA XX

PI Hill AVS, McShane H, Gilbert S, Reece W, Schneider J;

DR XX

XX WPI; 2002-394098/42.

XX PT Inducing CD4+ T-cell response against target antigen by administering a

PT composition comprising a source of CD4+ epitopes which is a

PT non-replicating or replication impaired recombinant poxvirus vector -

XX PS Example 1; Page 21; 50pp; English.

XX The present invention relates to a method of inducing a CD4+ T-cell

CC response against a target antigen, by administering two different

CC compositions comprising a source of CD4+ T-cell epitopes of the target

CC antigen, where the second composition further includes an epitope which

CC is same as the epitope of the first composition, where the source of the

CC epitopes for the compositions is a non-replicating or replication

CC impaired recombinant poxvirus vector. The methods are useful for inducing

CC CD4+ T-cell immune responses against diseases such as tuberculosis, human

CC immunodeficiency virus (HIV) (persistent viral infection), malaria,

CC *Helicobacter pylori*, influenza, hepatitis (chronic hepatitis B and C),

CC cytomegalovirus (CMV), viral infection, herpes virus-induced disease,

CC leprosy and diseases caused by non-malarial protozoan parasite such as

CC Mycobacterium tuberculosis useful in the method of the invention.

CC Sequence 15 AA;

SQ Query Match 100.0%; Score 79; DB 23; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.2e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTEQQWNFAGIEAA 15

Db 1 MTEQQWNFAGIEAA, 15

RESULT 3

AAY88581

ID AAY88581 standard; peptide: 17 AA.

XX AAY88581;

AC XX

DT 11-AUG-2000 (first entry)

XX DE Antigenic N-terminal sequence of ESAT-6.

XX KW ligand presenting assembly; early secreted antigen target 6; ESAT-6; bacteria; detect; diagnosis; allergen; cancer; vaccine; immune response; neurotropic factor; autoimmune system related compound; LPA; fungi; parasite; cell-adhesion molecule.

XX OS *Mycobacterium tuberculosis*.

XX FH Key

XX Location/qualifiers

XX 17 /note= "Ala at position 17 is linked via a linker to Ala

XX at position 17 of an identical peptide, where the linker is COCH2-CH(NH-Lys)-CH2CO, and the Lys residue contained in the linker is also linked to another peptide (see AAY88579);"

XX PN WO200018791-A1.

XX PD 06-APR-2000.

XX PD XX

PF 29-SEP-1999; 99WO-DR00510.

PF XX

PR 29-SEP-1998; 98DK-0001233.

PR XX

PA (STAT-) STAVENS SERUM INST.

PA XX

(HOLM/ ) HOLM A.

PA XX

PI Holm A, Jorgensen RM, Ostergaard S, Theisen M;

XX DR XX

XX DR WPI; 2000-303438/26.

XX PT

New ligand presenting assemblies useful for diagnosis, treatment and prevention of diseases caused by e.g. viruses, bacteria, toxins, allergens, autoimmune system-related compounds, cancer-related compounds, cell adhesion molecules -

XX XX

PT New ligand presenting assemblies useful for diagnosis, treatment and prevention of diseases caused by e.g. viruses, bacteria, toxins, allergens, autoimmune system-related compounds, cancer-related compounds, cell adhesion molecules -

XX XX

PT This sequence represents an antigenic N-terminal sequence from the

CC *Mycobacterium tuberculosis* early secreted antigenic target 6 (EST-6)

CC protein. The peptide is presented on the new ligand presenting assembly

CC of the invention. The invention relates to method for preparing ligand

CC presenting assemblies (LPAs) comprising:

CC (a) providing by solid phase synthesis, or fragment coupling, ligands

CC comprising desired sequences (e.g. the present sequence), the ligands

CC being attached to a solid phase;

CC (b) if necessary, deprotecting any N-terminal amino groups while the

CC ligands are still attached to the solid phase;

CC (c) reacting the ligands having unprotected N-terminal amino groups with

CC an achiral di-, tri- or tetracarboxylic acid, to provide a construct

CC having a ring structure; and

CC	(d) clearing the construct from the solid phase, to provide an LPA comprising ligands having tree C-terminal groups. The LPAs can be used for raising an immune response in an animal. They can also be used in vaccines and for generating antibodies in an animal. Alternatively they can be used for the treatment, alleviation, detection, diagnosis, or prophylaxis of diseases caused by viruses, bacteria, toxins, allergens, autoimmune system-related compounds, cancer related compounds, cell adhesion molecules, neurotropic factors, fungi or parasites. Use of the method enables the preparation of very long ring systems interconnected by reaction with the achiral di-, tri- or tetracarboxylic acid. The ring structure formed between desired sequences further enables additional presentation of desired sequences and chemical moieties. The LPAs provide very flexible systems for poly-functional constructs, and furthermore, products of high purity are obtained.
CC	Sequence 17 AA;
CC	Query Match 10.0%; Score 79; DB 21; Length 17;
CC	Best Local Similarity 10.0%; Pred. No. 1.3e-06; Mismatches 0; Indels 0; Gaps 0
QY	1 MTEQQWNFAGEEAA 15 
Db	1 MTEQQWNFAGEEAA 15
XX	
RESULT 4	
AAW35548	
ID AAW35548 standard; peptide; 19 AA.	
XX	
AC AAW35548;	
XX	
DT 22-APR-1998 (first entry)	
DE ESAT-6 antigen SEQ ID NO:120 from W09738011.	
XX	
KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour; scaffold; inhibition; metastasis; wound healing; solid phase.	
XX	
OS Synthetic.	
XX	
PN W09738011-A1.	
XX	
PD 16-OCT-1997.	
XX	
PF 03-APR-1997; 97WO-DE00146.	
XX	
PR 03-APR-1996; 96DK-0000398.	
XX	
PA (PEPR-) PEPRESEARCH AS.	
XX	
PI Heegaard PMH, Jakobsen PH;	
XX	
DR WPI: 1997-512645/47.	
XX	
PT Non-dendritic peptide carrier linked to a solid phase - useful as a diagnostic agent and as a scaffold for production of chemical derivatives	
PT	
XX	
PS Example 32; Page 156; 262pp; English.	
CC	A non-dendritic peptide carrier (A) has been developed which is coupled through a linker to a solid phase, forming a complex (A)-solid phase. Where (A) comprises 10-50 amino acids capable of forming a secondary structure in a benign buffer after liberation from the solid phase, and further the (A)-solid phase complex comprises an immunogenic substance and/or an immune mediator coupled on (A). The present sequence represents a peptide used in an example from the present invention. An (A)-solid phase complex can be used as a scaffold for the production of chemical derivatives, characterised by covalently attaching molecules at attachment points. Alternatively (A) is used as a scaffold-peptide for the incorporation into an immunostimulating Complex (Iscom) resulting from the incorporation of antigenic (A) Iscom complex which is used for the chemical coupling of antigenic

SQ	Sequence	19 AA:	CC	substances in an aqueous solution by conjugation. (A) derivatised with one or more Peptides having fibronectin-, laminin- or vitronectin-like binding activities can be used for the promotion of cell-attachment to plastic surfaces, in particular to inhibit tumour growth and metastasis, and for promotion of wound healing. Also a derivatised (A) can be used for the selection of specifically-binding aptamers or as a diagnostic agent. Such diagnostic-(A) molecules could be used to detect molecules derived from or indicative of pregnancy or of a disease, such as an infectious, autoimmune or cancerous disease.	
Query	Match	100.0%; Score 79; DB 18; Length 19;	CC		
Qy	Best Local Similarity	100.0%; Pred. No. 1.5e-06;	CC		
Matches	15; Conservative	0; Mismatches	CC		
Db	1 MTEQQWNFAGIEAA 15	Indels	0;	Gaps	
	1 MTEQQWNFAGIEAA 15				
RESULT 5	AAW32466	100.0%; Score 79; DB 18; Length 19;	CC		
ID	AAW32466 standard; Protein; 51 AA..	Best Local Similarity	100.0%; Pred. No. 1.5e-06;	CC	
XX	XX	Matches	15; Conservative	0; Mismatches	CC
AC	AAW32466;	Indels	0;	Gaps	CC
XX	09-JAN-1998 (first entry)				CC
DT					CC
XX	Mycobacterium tuberculosis antigen ESAT-6.				CC
DE					CC
XX	KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;				CC
KW	skin testing; M.tuberculosis.				CC
OS	Mycobacterium tuberculosis.				CC
XX					CC
PN	W09709428-A2.				CC
XX					CC
PD	13-MAR-1997.				CC
XX					CC
PF	30-AUG-1998; 96WO-US14674.				CC
XX					CC
PR	12-JUL-1996; 96US-0680574.				CC
PR	01-SEP-1995; 95US-0523436.				CC
PR	22-SEP-1995; 95US-0533634.				CC
PR	22-MAR-1996; 96US-062084.				CC
PR	05-JUN-1996; 96US-0659683.				CC
XX					CC
PA	(CORI-) CORIXA CORP.				CC
XX					CC
PI	Campbell-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;				CC
PI	Twardzik DR, Vedvick TH;				CC
XX					CC
DR	WPI; 1997-192903/17.				CC
DR	N-PSDB; AAN91529.				CC
XX					CC
PT	New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are useful in vaccines for prevention or treatment of tuberculosis, also for diagnosis				CC
PT					CC
PS	Disclosure: Page 135; 168pp; English.				CC
XX					CC
CC	A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M.tuberculosis antigen, ESAT-6. The immunogenic protein, and fusion proteins containing one or more of the proteins or one of the proteins Plus ESAT-6, are useful in vaccines, preferably when formulated with a non-specific adjuvant, to induce an immune response against M. tuberculosis (for treatment or prevention).				CC
Sequence	51 AA;				CC

Query Match 100.0%; Score 79; DB 18; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEQQWNFAGIEAA 15  
 XX |||||||  
 Db 1 MTEQQWNFAGIEAA 15

---

RESULT 6  
 ID AAW32339  
 XX AAW32339 standard; Protein; 51 AA.  
 AC AAW32339;  
 XX  
 DT 13-JAN-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen ESAT-6.  
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M tuberculosis.  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9709429-A2.  
 XX  
 PD 13-MAR-1997.  
 XX  
 PF 30-AUG-1996; 96WO-US14675.  
 XX  
 PR 12-JUL-1996; 96US-0680573.  
 PR 01-SEP-1995; 95US-0523435.  
 PR 22-SEP-1995; 95US-0532136.  
 PR 22-MAR-1996; 96US-0620280.  
 PR 05-JUN-1996; 96US-0658800.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PR Campos-Neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PT Twardzik DR, Vedvick TH;  
 XX  
 DR WPI; 1997-192904/17.  
 DR N-PSDB; AAT91463.  
 XX  
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 - useful for diagnosis of M. tuberculosis infection  
 XX  
 PS Claim 43; Page 147; 190pp; English.  
 XX  
 CC A new immunogenic polypeptide has been developed comprising an  
 immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 its variant differing only in conservative substitutions and/or  
 modifications). The present sequence represents a specifically claimed  
 sequence of the ESAT-6 M. tuberculosis antigen. The immunogenic  
 polypeptide can be used to diagnose M. tuberculosis infection by forming  
 complexes with specific antibodies in the sample. Fragments of DNA  
 encoding the immunogenic polypeptide can be used as diagnostic primers  
 or probes and agents that bind to the antigen, especially monoclonal  
 antibodies or equivalent polyclonal antibodies, are also used for  
 CC diagnosis.  
 XX  
 Sequence 51 AA;

Query Match 100.0%; Score 79; DB 19; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEQQWNFAGIEAA 15  
 XX |||||||  
 Db 1 MTEQQWNFAGIEAA 15

---

RESULT 7  
 ID AAW81701  
 ID AAW81701 standard; Protein; 51 AA.  
 XX  
 AC AAW81701;  
 XX  
 DT 27-JAN-1999 (first entry)  
 XX  
 DE Mycobacterium immunogenic polypeptide ESAT-6.  
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9816646-R2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PR Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1998-261042/23.  
 DR N-PSDB; AAV64501.  
 XX  
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 PR Disclosure; Page 126; 230pp; English.  
 XX  
 CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 51 AA;

Query Match 100.0%; Score 79; DB 19; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEQQWNFAGIEAA 15  
 XX |||||||  
 Db 1 MTEQQWNFAGIEAA 15

---

RESULT 8  
 ID AAW64334  
 ID AAW64334 standard; Protein; 51 AA.  
 XX  
 AC AAW64334;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen ESAT-6.  
 XX  
 KW Tuberculosis; infection; diagnosis; antigen; ESAT-6.  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9816645-A2.  
 XX  
 PD 23-APR-1998.



XX Sequence 51 AA;  
 SQ 5Q

Query Match 100.0%; Score 79; DB 20; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQQWNFAGIEAAA 15  
 Db 1 MTEQQWNFAGIEAAA 15

RESULT 11  
 AAU1897 ID AAU1897 standard; Protein; 51 AA.  
 XX AC AAU1897;  
 XX DT 29-AUG-2001 (first entry)  
 DE Mycobacterium tuberculosis partial antigen ESAT-6.  
 KW ESAT-6; antigen; vaccine; tuberculosis; AIDS;  
 KW acquired immunodeficiency disease.  
 XX OS Mycobacterium tuberculosis.  
 XX PN WO200124820-A1.

XX PD 12-APR-2001.  
 PF 10-OCT-2000; 2000WO-US28095.  
 PR 07-OCT-1999; 99US-0158338.  
 PR 07-OCT-1999; 99US-0158425.  
 PA (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;  
 DR N-PSDB; AAS03787.

XX PT Vaccinating against Mycobacteria infections in mammals using fusion proteins comprising combinations of heterologous antigens -  
 PR WPI: 2001-290576/30.

XX Disclosure; Page 164; 168pp; English.

The sequence represents Mycobacterium tuberculosis ESAT-6 antigen. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by Mycobacteria. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with Mycobacterium tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS.

XX Sequence 51 AA;

SQ 5Q

Query Match 100.0%; Score 79; DB 22; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQQWNFAGIEAAA 15  
 Db 1 MTEQQWNFAGIEAAA 15

RESULT 13  
 AAW11494 ID AAW11494 standard; Protein; 95 AA.  
 XX AC AAW11494;  
 XX DT 27-MAR-1997 (first entry)

RESULT 12  
 RRESULT 12  
 ID AAE17580  
 ID AAE17580 standard; Protein; 51 AA.  
 XX AC AAE17580;

XX DT 22-APR-2002 (first entry)  
 XX DE Mycobacterium species ESAT-6 protein.  
 XX KW Fusion protein; antigen; serological sensitivity; immune response; tuberculosis; infection; vaccine; ESAT-6 protein.  
 OS Mycobacterium sp.  
 XX PN WO200198460-A2.

XX PD 27-DEC-2001.  
 XX PR 20-JUN-2001; 2001WO-US19959.  
 XX PR 01-FEB-2001; 2001US-205737P.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Skeiky Y, Reed S, Alderson M;  
 DR N-PSDB; AAD28351.  
 XX PT Composition comprising MTB39 antigen and MTB32A antigen from Mycobacterium species, useful for eliciting immune response in a subject -  
 PT subject -  
 XX PS Claim 9; Page 122; 13pp; English.  
 XX The present invention relates to fusion proteins containing at least two Mycobacterium species antigens, nucleotides encoding them and compositions comprising such fusion proteins. The present invention particularly relates to nucleic acids encoding fusion proteins that include two or more individual *M. tuberculosis* antigens which increase the serological sensitivity of sera from individuals infected with tuberculosis and methods for their use in diagnosis, prevention and treatment of tuberculosis infection. Sequences of the invention are useful for eliciting an immune response in a mammal, e.g., human, immunised with BCG. They are useful in the diagnosis, treatment and prevention of *Mycobacterium* infection. The fusion proteins and the polynucleotides are useful as diagnostic tools in patients infected with *Mycobacterium*, in vitro and in vivo assays for detecting humoral antibodies or cell-mediated immunity against *M. tuberculosis*, for the diagnosis of an infection or monitoring of disease progression, as immunogens to generate or elicit a protective immune response in a patient and for raising anti-*M. tuberculosis* antibodies in a non-human animal. Sequences of the invention are also used as vaccines. MTB32A fusion proteins of the invention are useful as in vivo diagnostic agents for intradermal skin test. The present sequence is Mycobacterium species ESAT-6 protein.  
 XX Sequence 51 AA;  
 SQ 5Q

Query Match 100.0%; Score 79; DB 23; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-06; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEQQWNFAGIEAAA 15  
 Db 1 MTEQQWNFAGIEAAA 15

DE ESAT6.  
 XX  
 KW ESAT6; Mycobacterium tuberculosis; 6kDa antigen; tuberculosis complex;  
 KW HYB6-8 reactive antigen; interferon-gamma; memory T-lymphocyte; vaccine;  
 KW M. africanum; M. bovis; delayed-type hypersensitivity reaction; therapy.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9501441-A1.  
 XX  
 PD 12-JAN-1995.  
 XX  
 PF 01-JUL-1994; 94W0-DK00273.  
 XX  
 PR 02-JUL-1993; 93DK-0000798.  
 XX  
 PA (STAT-) STATENS SERUMSINSTITUT.  
 XX  
 PI Andersen AB, Andersen P, Haslov K, Sorensen A;  
 XX  
 DR WPI; 1995-061005/08.  
 DR N-PSDB; AAT51422.  
 XX  
 Vaccine for tuberculosis induces interferon-gamma release from T-lymphocytes - comprises an antigen released from mycobacteria, for immunisation of humans  
 PT Claim 5; Page 61-63; 101pp; English.

XX  
 CC This sequence represents the Mycobacterium tuberculosis ESAT6, also known as the 6kDa antigen, or the HYB6-8 reactive antigen. ESAT6 is released from metabolising bacteria, and can be isolated from short term filtrates grown as shaken cultures for 7 days. ESAT6 also induces a release of interferon-gamma from reactivated memory T-lymphocytes. This protein sequence is included in the vaccine of the invention. The vaccine is for immunising an animal (including humans) against tuberculosis caused by a Mycobacteria belonging to the tuberculosis complex. The Mycobacteria of the tuberculosis complex are, M. tuberculosis, M. africanum, and M. bovis. The vaccine can evoke a protective immune response against tuberculosis or a delayed-type hypersensitivity reaction. The protein can also be included in a composition for diagnosing tuberculosis. The composition is injected intradermally, and a skin reaction is an indicator of tuberculosis.  
 XX  
 Sequence 95 AA;

Query Match 100.0%; Score 79; DB 16; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MTEQQNMFAGIEAA 15  
 Db 1 MTEQQNMFAGIEAA 15

SQ Sequence 95 AA;

Query Match 100.0%; Score 79; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MTEQQNMFAGIEAA 15  
 Db 1 MTEQQNMFAGIEAA 15

RESULT 14  
 AAY29890  
 ID AAY29890 standard; Protein; 95 AA.  
 XX  
 AC AAY29890;  
 XX  
 DT 18-Nov-1999 (first entry)  
 DE Mycobacterium tuberculosis ESAT-6 protein sequence.  
 XX  
 KW Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;  
 KW delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;  
 KW interferon-gamma release.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9945119-A2.  
 XX

RESULT 15  
 AAY29788  
 ID AAY29788 standard; Protein; 95 AA.  
 XX  
 AC AAY29788;  
 XX  
 DT 08-Nov-1999 (first entry)  
 DE Mycobacterial tuberculosis ESAT-6 protein.  
 XX  
 KW Mycobacterial; ESAT-6; 6 kDa antigen; tuberculosis; infection;  
 KW immunological response; diagnosis; vaccine.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN US5955077-A.  
 XX  
 PD 21-SEP-1999.

XX  
 PF 05-JUN-1995; 95US-0465640.  
 XX PR 05-JUN-1995; 95US-0465640.  
 PR 20-SEP-1993; 93US-0123182.  
 PR 01-JUL-1994; 94WO-DK00273.  
 XX PA (STAT-) SERUMINSTITUT.  
 XX Andersen AB, Andersen P, Haslov K, Sorensen AL;  
 XX DR WPI; 1999-539545/45.  
 N-PSDB; AA208877.

PT Polypeptide secreted from *Mycobacterium* is useful as a vaccine  
 PT against tuberculosis

PS XX Claim 24; FIG 10C; 39pp; English.

CC The present invention describes a purified or non-naturally occurring  
 CC polypeptide (I) released from a metabolising mycobacteria comprising an  
 CC ESAT6 also called the 6 kDa antigen. The present sequence represents  
 CC *Mycobacterium tuberculosis* ESAT-6 protein. Also described is a purified  
 CC or non-naturally occurring Polypeptide (II) with the ability to elicit  
 CC a delayed type hypersensitivity reaction which comprises a T cell  
 CC epitope of (I). (II) can be used with a carrier or vehicle in a  
 CC composition for diagnosing tuberculosis caused by mycobacteria belonging  
 CC to the tuberculosis complex, i.e. *Mycobacterium tuberculosis*, *M. bovis*  
 CC and *M. africanum*. The composition can be used to detect microbial  
 CC antibodies or components of mycobacteria in samples or in animals  
 CC through the use of immunoassays. (III) can be used as a vaccine for  
 CC immunising an animal, including humans against tuberculosis caused by  
 CC mycobacteria of the tuberculosis-complex. (II) induce a release of  
 CC IFN-gamma from reactivated T-lymphocytes evoking a protective immune  
 CC response. Vaccine containing (II) has the same protective potency as  
 CC the live BCG vaccine against tuberculosis.

XX Sequence 95 AA;

SO Query Match 100.0%; Score 79; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 8 1e-06; Mismatches 0;  
 Matches 15; Conservative 0; Indels 0; Caps 0;

Qy	1	MTEQOWNFAGIEAA 15
	1	
Db	1	MTEQOWNFAGIEAA 15

Search completed: July 3, 2003, 14:05:31  
 Job time : 31. secs

GenCore version 5.1.6  
copyright (c) 1993 - 2003 Compugen Ltd.

### OM protein - protein search, using sw model

Run on: July 3, 2003, 14:02:14 ; Search time 6.5 Seconds

(without alignments)  
95.715 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: MTEQQWNFAGIEAA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No. Score Query Match Length DB ID

### DESCRIPTION

Result No.	Score	Query	Match	Length	DB	ID	Description
1	74	93.7	94	1	ESAG_MYCTU	Q57165	mycobacteri
2	57.0	57.0	476	1	NOREA_RHIME	052892	rhizobacteri
3	42	53.2	539	1	G6PL_CAURR	Q9abt5	caulobacter
4	41	51.9	176	1	DSBB_ECOLI	P30018	escherichia
5	41	51.9	176	1	DSBB_SHIFL	054155	shigella fl
6	38	48.1	305	1	YN45_DETRA	Q9rrv7	deinococcus
7	38	48.1	438	1	SGCE_HUMAN	043556	homo sapien
8	38	48.1	444	1	CADB_ECOLI	P23891	escherichia
9	38	48.1	508	1	HNCM_PIG	Q02734	sus scrofa
10	38	48.1	652	1	DREB2_CHICK	P18302	gallus gall
11	38	48.1	2564	1	SPCQ_HUMAN	Q9h254	homo sapien
12	37	46.8	211	1	MTO4_SYNTH3	P73161	synthecyst
13	37	46.8	264	1	COMT_RAT	P22734	rattus norv
14	37	46.8	294	1	G3P_ESCBLL	P24149	escherichia
15	37	46.8	335	1	YD59_SCHPO	Q10312	schizosacch
16	37	45.8	354	1	WNL_BRARE	Q73864	brachydorina
17	37	45.8	445	1	YJDE_ECOLI	P39269	escherichia
18	37	46.8	475	1	GUNA_CLOCHE	P17901	clostridium
19	37	46.8	668	1	YNL8_BEAST	Q04511	saccharomy
20	37	46.8	879	1	SYL_XYLIFA	Q9pbq8	xylella fast
21	36.5	46.2	947	1	LKTA_PASSP	P55223	pasteurella
22	36	45.6	131	1	INBN_SHITY	P03832	shigella dy
23	36	45.6	148	1	HPCR_ECOLI	Q07195	escherichia
24	36	45.6	175	1	RP18_RCAT	P47940	rattus norv
25	36	45.6	230	1	TERM_RCP1	Q37988	bacterioph
26	36	45.6	326	1	ODPB_RICP1	Q9zdr3	licketticia
27	36	45.6	430	1	PYRC_LACPL	P77884	lactobacilli
28	36	45.6	475	1	ARCD_LACSK	053092	lactobacilli
29	36	45.6	496	1	SRM_MOUSE	Q62270	mus musculu
30	36	45.6	499	1	HURE_BRUME	Q8Yi71	brucella me
31	36	45.6	511	1	HUTR_RHIME	Q31197	rhizobium m
32	36	45.6	700	1	EFG2_RALSO	Q8xrm1	raistonia s
33	36	45.6	901	1	SBCA_HAEIN	P43803	haemophilus

### ALIGNMENTS

RESULT 1	ESA6_MYCTU	ID	ESA6 MYCTU	STANDARD;	PRT;	94 AA.
		AC	Q57165; 084901;			
		DT	30-MAY-2000 (Rel. 39, Last sequence update)			
		DT	15-JUN-2002 (Rel. 41, Last annotation update)			
		DE	6 kDa early secretory antigenic target (ESAT-6).			
		GN	ESAT6 OR RV3875 OR MT3989 OR MTv027.10.			
		OS	Mycobacterium tuberculosis, and			
		OC	Bacteria; Actinobacteria; Corynebacterineae; Mycobacteriaceae; Actinomycetales; Corynebacterineae; Mycobacterium.			
		OX	NCBI_TAXID=1773, 1765;			
		RN	[1]			
		RP	SEQUENCE FROM N.A.			
		RC	SPECIES=M_tuberculosis; STRAIN=H37Rv;			
		RX	MEDLINE=95204331; PubMed=7897219;			
		RA	Andersen P., Andersen A.B., Sorensen A.L., Nagai S.:			
		RA	"Recall of long-lived immunity to <i>Mycobacterium tuberculosis</i> infection in mice." J. Immunol. 154:3359-3372(1995).			
		RL	J. Immunol. 154:3359-3372(1995).			
		RN	[2]			
		RP	SEQUENCE FROM N.A.; SEQUENCE OF 1-10, AND CHARACTERIZATION.			
		RC	SPECIES=M_tuberculosis; STRAIN=ETGmann;			
		RX	MEDLINE=95247251; PubMed=729876;			
		RA	Soerensen A.L., Nasai S., Hogen G., Anderson P., Anderson A.B.;			
		RA	"Purification and characterization of a low-molecular-mass T-cell antigen secreted by <i>Mycobacterium tuberculosis</i> ."; Infect. Immun. 63:1710-1717(1995).			
		RA	[3]			
		RP	SEQUENCE FROM N.A.			
		RC	SPECIES=M_tuberculosis; STRAIN=H37Rv;			
		RX	MEDLINE=98295887; PubMed=964230;			
		RA	Cola S.T., Brisch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Elsegmeyer K., Gas S., Barry C.E. III, Telaiola F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hoiryd S., Hornsby T., Jangels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Rajandream M.A., Rogers J., Rutler S., Seeger K., Shelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.; "Deciphering the biology of <i>Mycobacterium tuberculosis</i> from the complete genome sequence"; Nature 393:537-544(1998).			
		RA				
		RN	[4]			
		RP	SEQUENCE FROM N.A.			
		RC	SPECIES=M_tuberculosis; STRAIN=CDC 1551 / Oshkosh;			
		RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Brinkman F., Salzberg S.L., Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;			
		RA	"Whole genome comparison of <i>Mycobacterium tuberculosis</i> clinical and laboratory strains"; Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.			

RN [5]  
 RP SQUENCE FROM N.A.  
 RC SPECIES=M. tuberculosis;  
 RA Singh B., Siddiqui Z., Singh S., Sharma P.;  
 RT "ESAT-5 gene of a clinical isolate of *Mycobacterium tuberculosis* from  
 RA India.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SQUENCE FROM N.A.  
 RC SPECIES=M. bovis;  
 RX MEDLINE=9620095; PubMed=8631702;  
 RA Mahairas G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;  
 RT "Molecular analysis of genetic differences between *Mycobacterium bovis*  
 BCG and virulent M. bovis";  
 RL J. Bacteriol. 178:1274-1282(1996).  
 RN [7]  
 RP SQUENCE OF 1-70 FROM N.A.  
 RC SPECIES=M. tuberculosis; STRAIN=H37Rv;  
 RX MEDLINE=99061212; PubMed=9446755;  
 RA Berthet F.-X., Rasmussen P.B., Rosenkranz I., Andersen P.,  
 RA Glcquel B.;  
 RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel  
 low-molecular-mass culture filtrate protein (cPP-10)." ;  
 RL Microbiology 144:3195-3203(1998).  
 CC -1- FUNCTION: NOT KNOWN; HIGH LEVEL OF INF-GAMMA FROM MEMORY  
 CC EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE  
 CC RESPONSE.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial/  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

---

CC -1- FUNCTION: NOT KNOWN; DOES NOT SEEM TO PARTICIPATE IN NOD FACTOR  
 CC SYNTHESIS BUT REQUIRED FOR NODULATION ON SOME SPECIFIC MEDICAGO  
 CC SPECIES SUCH AS M. LITTORALIS.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial/  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

---

CC -1- FUNCTION: NOT KNOWN; DOES NOT SEEM TO PARTICIPATE IN NOD FACTOR  
 CC SYNTHESIS BUT REQUIRED FOR NODULATION ON SOME SPECIFIC MEDICAGO  
 CC SPECIES SUCH AS M. LITTORALIS.

DR U34848 AAC44033.1;-  
 DR EMBL: X79562; CAA16090.1;-  
 DR EMBL: AA022120; CAA17967.1;-  
 DR EMBL: AF420491; AAC16896.1;-  
 DR EMBL: AE007190; ANK48357.1;-  
 DR EMBL: AF004671; AAC83446.1;-  
 DR TIGR: MT3989;-  
 DR Tuberculist: Rv3875;-  
 KW Antigen; Complete proteome.  
 FT INIT\_MER 0  
 FT END\_MER 0  
 SQ SQUENCE 94 AA; 973 MW; 19245B0EC478BC84 CRC64;

Query Match 93.7%; Score 74; DB 1; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 QWNFGIERRA 15  
 Db 100 ENSFAGLKAA 110

RESULT 2  
 NOEA\_RIME STANDARD; PRM; 476 AA.  
 ID NOEA\_RIME  
 AC 052892;  
 DT 30-MAY-2000 (Rel: 39, Created)  
 DT 30-MAY-2000 (Rel: 39, Last sequence update)  
 DT 15-JUN-2002 (Rel: 41, Last annotation update)

OY 2 TEQDNFFAGIEAA 15  
 1 TEQDNFFAGIEAA 14

RESULT 3  
 G6PI\_CAUCR STANDARD; PRT; 539 AA.  
 ID G6PI\_CAUCR  
 AC QABK5;  
 DT 15-JUN-2002 (Rel: 41, Created)  
 DT 15-JUN-2002 (Rel: 41, Last sequence update)  
 DT 15-JUN-2002 (Rel: 41, Last annotation update)  
 DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphogluucose  
 DE isomerase) (PGI) (Phosphohexose isomerase) (PHI).  
 GN PG1 OR CC0222.  
 OS Caulobacter crescentus  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter;  
 OX NCBI\_TAXID=155892;  
 RN [1]  
 RP SQUENCE FROM N.A.  
 RC STRAIN=ATCC-19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyst T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka T., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermlaeva M., White O.,  
 RA NCBI\_TAXID=382;

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RA "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001);  
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-  
 phosphate.  
 CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.  
 CC -1- SUBCELLULAR LOCATION: Cyttoplasmic (By similarity).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 EMBL: AEG05696; ARK22209\_1; -  
 DR HSSP; Q8NLE2; 1H0X.  
 DR TIGR; CC0222; -  
 DR InterPro; IPR001672; G6P\_Isomerase.  
 DR Pfam; PF0042; PG1; 1.  
 DR PRINTS; PR00562; G6PISOMERASE.  
 DR PROSITE; PS00765; P\_GLUCOSE\_ISOMERASE\_1; 1.  
 DR PROSITE; PS00174; P\_GLUCOSE\_ISOMERASE\_2; 1.  
 KW Isomerase; Gluconogenesis; Glycolysis; Complete proteome.  
 FT ACT\_SITE 380 508 BY SIMILARITY.  
 AC SITE 508 BY SIMILARITY.  
 SEQUENCE 539 AA: 57526 MW: B233DFRAF7FDI595 CRC64;  
 Query Match 53.2%; Score 42; DB 1; Length 539;  
 Best Local Similarity 50.0%; Pred. No. 9.6;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MTEQWNNPAGIEAA 14  
 ::::: 1::1::1|  
 Db 48 LSKQAWDANGLEAA 61  
 RESULT 4  
 DSBB\_ECOLI STANDARD; PRT: 176 AA.  
 AC P30018;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE disulfide bond formation protein B (Disulfide oxidoreductase).  
 GN DSBB OR ROXB OR BL185 OR Z1448 OR EC51680.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia; NCBL\_TaxID=562, 83334;  
 RN [1] SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RX MEDLINE-93157338; PubMed=840071;  
 RA Barcwell J.C.A., Lee J.-O., Jander G., Martin N., Belin D.,  
 RA Beckwith J.;  
 RT "A pathway for disulfide bond formation in vivo";  
 RT Proc. Natl. Acad. Sci. U.S.A. 90:1038-1042(1993);  
 RN [2] SEQUENCE FROM N.A.  
 RX MEDLINE-9334817; PubMed=7688471;  
 RA Missakas D., Georgopoulos C., Reina S.;  
 RT "Identification and characterization of the Escherichia coli gene dsbb, whose product is involved in the formation of disulfide bonds in vivo";  
 RT Proc. Natl. Acad. Sci. U.S.A. 90:7084-7088(1993);  
 RN [3] SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Geelen M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC [4] SEQUENCE FROM N.A.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97061202; PubMed=8905232;  
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kojihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Maki T., Micobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nasimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seiki Y., Tagami H., Takekoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horuchi T.,  
 RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [5] SEQUENCE FROM N.A.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / ATCC 700927;  
 RX MEDLINE-21074335; PubMed=11206551;  
 RA Perera N.T., Prunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rosa D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamitis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Battner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli 0157:H7.";  
 RT Nature 409:529-533(2001).  
 RN [6] SEQUENCE FROM N.A.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RTMD 0509952;  
 RX MEDLINE-21156331; PubMed=1158796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takani H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli 0157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [7] PRELIMINARY SEQUENCE OF 1-169 FROM N.A.  
 RP PRELIMINARY SEQUENCE OF 1-169 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE-9228303; PubMed=1317851;  
 RA Pineda E., Padan E., Schulding S.;  
 RT "Cloning, sequencing, and expression of the nhab gene, encoding a Na<sup>+</sup>/H<sup>+</sup> antiporter in Escherichia coli";  
 RL J. Biol. Chem. 267:11064-11068(1992).  
 RN [8] TOPOLOGY, AND MUTAGENESIS OF CYSTEINE RESIDUES.  
 RX MEDLINE-95045404; PubMed=7957076;  
 RA Jander G., Martin N.L., Beckwith J.;  
 RT "Two cysteines in each periplasmic domain of the membrane protein Dsbb are required for its function in protein disulfide bond formation.";  
 RL EMBO J. 13:5121-5127(1994).  
 RN [9] ACTIVE SITES.  
 RP MEDLINE-99164086; PubMed=10064586;  
 RA Kobayashi T., Ito K.;  
 RT "Respiratory chain strongly oxidizes the CXXC motif of Dsbb in the Escherichia coli disulfide bond formation pathway.";  
 RL EMBO J. 18:1192-1198(1999).  
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME PERIPLASMIC PROTEINS SUCH AS PHOA OR OMFA. ACTS BY OXIDIZING THE DSB PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE DSBB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; LQ3721; AAA23711.1; ALT\_INIT.

CC DR EMBL; AE000216; AAC71269.1; -.

CC DR EMBL; D9752; BAA35032.1; ALT\_INIT.

CC DR EMBL; D90753; BAA35040.1; ALT\_INIT.

CC DR EMBL; AE000335; AAG55036.1; -.

CC DR EMBL; AP00255; BAB35103.1; -.

CC DR EMBL; M83655; -; NOT\_ANNOTATED\_CDS.

CC PIR; JCL109; JCL110.

CC DR EcoGene; EG11393; dsbB.

CC DR InterPro; IPR003752; DsbB.

CC DR Pfam; PF02600; DsbB; 1.

CC DR Oxidoreductase; Redox-active center; Electron transport; Chaperone; Transmembrane; Inner membrane.

CC KW Redox-active center; Electron transport; Chaperone; Transmembrane; Inner membrane; Complete proteome; Cyttoplasmic (PROBABLE).

CC FT DOMAIN 1 14 PROBABLE.

CC FT TRANSMEM 15 31 PROBABLE.

CC FT DOMAIN 32 49 PROBABLE.

CC FT TRANSMEM 50 65 PROBABLE.

CC FT DOMAIN 66 71 PROBABLE.

CC FT TRANSMEM 72 89 PROBABLE.

CC FT DOMAIN 90 144 PROBABLE.

CC FT TRANSMEM 145 162 PROBABLE.

CC FT DOMAIN 163 176 PROBABLE.

CC FT DISULFID 41 44 PROBABLE.

CC FT DISULFID 104 130 PROBABLE.

CC FT CONFLICT 50 51 AL > GLG (IN REF. 8).

CC FT SEQUENCE 136 136 D -> DL (IN REF. 8); 20142 MW; 9GBD673D51E9F09B CRC64;

CC Query Match 51.9%; Score 41; DB 1; Length 176; Best Local Similarity 60.0%; Pred. No. 4.7; Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CC QY 3 EDOWNFAE 12 1:1:1 1:1

CC Db 132 EROWDFGLE 141

RESULT 5

DSBB\_SHIFL

ID DSBB\_SHIFL STANDARD; PRT; 176 AA.

AC 054155; -.

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Disulfide bond formation protein B (Disulfide oxidoreductase).

GN DSBB.

OS Shigella flexneri.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Shigellaceae; NCBI\_TAXID=623; NCBI\_TAXID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R1;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D., Dodson R.J., Haft D.H., Grinn M.L., Nelson W.C., Richardson D.L., Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M., Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C., Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D., Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C., Fraser C.M., RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1", 157(1999), 286-1571.

RL Science 286:1571-1577(1999).

CC -1-SIMILARITY: BELONGS TO THE UPF0151 FAMILY. SOME SIMILARITY TO STRAIN=YSH6000 / serotype 2A; Sasakawa C.;

CC Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

CC -1-FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME PERIPLASMIC PROTEINS SUCH AS PHO OR OMPA. ACTS BY OXIDIZING THE DBA PROTEIN (BY SIMILARITY).

CC -1-SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (By similarity).

CC -1-SIMILARITY: BELONGS TO THE DSBB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC DR EMBL; D38244; BAA07408.1; -.

CC DR InterPro; IPR003752; DsbB.

CC DR Pfam; PF02600; DsbB; 1.

CC DR Oxidoreductase; Redox-active center; Electron transport; Chaperone; Transmembrane; Inner membrane.

CC FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).

CC FT TRANSMEM 15 31 CYTOPLASMIC (POTENTIAL).

CC FT DOMAIN 32 49 CYTOPLASMIC (POTENTIAL).

CC FT TRANSMEM 50 65 CYTOPLASMIC (POTENTIAL).

CC FT DOMAIN 66 71 CYTOPLASMIC (POTENTIAL).

CC FT TRANSMEM 72 89 CYTOPLASMIC (POTENTIAL).

CC FT DOMAIN 90 144 CYTOPLASMIC (POTENTIAL).

CC FT TRANSMEM 145 162 CYTOPLASMIC (POTENTIAL).

CC FT DOMAIN 163 176 CYTOPLASMIC (POTENTIAL).

CC FT DISULFID 41 44 REDOX-ACTIVE (BY SIMILARITY).

CC FT DISULFID 104 130 REDOX-ACTIVE (BY SIMILARITY).

CC SQ SEQUENCE 176 AA; 20031 MW; 7552F2B916B72246 CRC64;

CC Query Match 51.9%; Score 41; DB 1; Length 176; Best Local Similarity 60.0%; Pred. No. 4.7; Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CC QY 3 EDOWNFAE 12 1:1:1 1:1

CC Db 132 EROWDFGLE 141

RESULT 6

VN45\_DMRA

ID VN45\_DMRA STANDARD; PRT; 305 AA.

AC 09R77; DEBRA

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein DR2345.

GN DR2345.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales; Deinococcaceae; Deinococcus.

OC Deinococcaceae; Deinococcus.

OK NCBI\_TAXID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R1;

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D., Dodson R.J., Haft D.H., Grinn M.L., Nelson W.C., Richardson D.L., Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M., Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C., Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D., Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C., Fraser C.M., RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1", 157(1999), 286-1571.

RL Science 286:1571-1577(1999).

CC -1-SIMILARITY: BELONGS TO THE UPF0151 FAMILY. SOME SIMILARITY TO STRAIN=YSH6000 / serotype 2A; Sasakawa C.;

CC Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

CC -1-FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME PERIPLASMIC PROTEINS SUCH AS PHO OR OMPA. ACTS BY OXIDIZING THE DBA PROTEIN (BY SIMILARITY).

CC -1-SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (By similarity).

CC -1-SIMILARITY: BELONGS TO THE DSBB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC DR EMBL; AE00265; AAF11890.1; -.

CC DR TIGR; DR2345; -.

CC DR InterPro; IPR004843; M-Peptidase.

CC DR InterPro; IPR004844; S/T\_phosphotse.

CC DR Pfam; PF00149; Metallophos; 1.

CC DR Hypothetical protein; Hydrolase; Complete proteome.



RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12";  
 RN DNA Res. 8:11-22(2001).  
 RL -1 - FUNCTION: PROBABLE CADAVERINE/LYSINE ANTIPORTER OR PAPR OR IT.  
 CC -1 - SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1 - INDUCTION: AT LOW PH IN THE PRESENCE OF LYSINE AND ACHIEVES  
 CC MAXIMAL LEVEL UNDER ANAEROBIC CONDITIONS.  
 CC -1 - SIMILARITY: BELONGS TO THE ARCD/CADB/GERAB/LYSI FAMILY OF  
 CC PERMEASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; MG7452; AAA23532; 1; -  
 DR EMBL; M76411; AA23535; 1; -  
 DR EMBL; 014003; AAAX7032; 1; -  
 DR EMBL; AB000486; AAC77093; 1; -  
 DR EMBL; AE005647; ANG5932; 1; -  
 DR EMBL; AP002568; BAB38537; 1; -  
 DR PIR; B41968;  
 DR PIR; A41842;  
 DR Ecogene; EG10132; cadB.  
 DR InterPro; IPR02223; AA\rel\\_Pmeasel.  
 DR InterPro; IPR00481; permease.  
 DR InterPro; IPR04154; R\O\\_Antipor.  
 DR Pfam; PF00324; aa\\_Permease; 1.  
 DR TIGRFAMS; TIGR00905; 2A0203; 1.  
 KW Transport; Antipor; Amino-acid transport; Transmembrane;  
 KW Inner membrane; Complete proteome.  
 FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 68 88 POTENTIAL.  
 FT TRANSMEM 95 115 POTENTIAL.  
 FT TRANSMEM 123 143 POTENTIAL.  
 FT TRANSMEM 149 169 POTENTIAL.  
 FT TRANSMEM 193 213 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 250 270 POTENTIAL.  
 FT TRANSMEM 273 293 POTENTIAL.  
 FT TRANSMEM 323 343 POTENTIAL.  
 FT TRANSMEM 354 374 POTENTIAL.  
 FT TRANSMEM 384 404 POTENTIAL.  
 FT TRANSMEM 405 425 POTENTIAL.  
 SQ SEQUENCE 444 AA; 46665 MW; E87913B449B0500A CRC64;  
 QY Query Match 48.1%; Score 38; DB 1; Length 444;  
 QY Best Local Similarity 60.0%; Pred. No. 38;  
 QY Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY  
 Db 198 WNFAGIEAA 15  
 Db 198 WNFAGIEAA 207

RESULT 9  
 HNCM\_PIG STANDARD; PRT; 508 AA.  
 ID HNCM\_PIG  
 AC 002734; 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor  
 DE (EC 4.1.3.5) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A  
 DE synthase).  
 GN HNC52.

OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NEBI\_TAXID=9823;  
 RN [1]  
 RP SOURCE FROM N.A. (ISOFORMS E1 AND E2).  
 RC TISSUE=Brain;  
 RX MEDLINE=89059137; PubMed=9164842;  
 RA Kohlma N., Kato Y., Shiao T., Obata K.;  
 RT "Gene expression of mitochondrial 3'-hydroxy-3'-methylglutaryl-CoA  
 synthase in a poorly ketogenic mammal: effect of starvation during  
 RT the neonatal period of the piglet.";  
 RL Tissue-Biochem. J. 324:65-73(1997).  
 CC MEDLINE=97233201; PubMed=9164842;  
 CC Adams S.H., Allo C.S., Asins G., Heggardt F.G., Marrero P.F.;  
 CC "Gene expression of mitochondrial 3'-hydroxy-3'-methylglutaryl-CoA  
 CC synthase in a poorly ketogenic mammal: effect of starvation during  
 CC the neonatal period of the piglet.";  
 CC  
 CC -1 - FUNCTION: THIS ENZYME CONDENSES ACETYL-COA WITH ACETOACETYL-COA  
 CC TO FORM HMG-COA, WHICH IS THE SUBSTRATE FOR HMG-COA REDUCTASE.  
 CC -1 - CATALYTIC ACTIVITY: (S)-3-hydroxy-3'-methylglutaryl-CoA + CoA =  
 CC acetyl-CoA + (H<sub>2</sub>O) + acetoadetyl-CoA.  
 CC -1 - PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE  
 CC BODY BIOSYNTHESIS.  
 CC -1 - SIMILARITY: BELONGS TO THE HMG-COA SYNTHASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; 090884; AAC48727; 1; -  
 DR InterPro; IPR00590; HMG\_Coa\_Synth.  
 DR Pfam; PF01154; HMG\_Coa\_Synth; 1.  
 DR PROSITE; PS01226; HMG\_COA\_SYNTHASE; 1.  
 KW Lyase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;  
 KW Multigene family.  
 FT TRANSIT 1 37 MITOCHONDRION (PROBABLE).  
 FT CHAIN 38 HYDROXYMETHYGLUTARYL-COA SYNTHASE.  
 FT ACT-SITE 166 POTENTIAL.  
 SQ SEQUENCE 508 AA; 56933 MW; 5479DB6F70B3C0F6 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 508;  
 Best Local Similarity 70.0%; Pred. No. 44;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 EQQDNFAGIE 12  
 QY 1:||| |||||  
 Db 278 EKOWKQAGIE 287

RESULT 10  
 DRB\_CHICK STANDARD; PRT; 652 AA.  
 ID DRB\_CHICK  
 AC P18302; 091358; 091359;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Drebrin (Developmentally regulated brain protein).  
 GN DN1  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauvia; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OC NEBI\_TAXID=9031;  
 RP  
 RN [1]  
 SQ SOURCE FROM N.A. (ISOFORMS E1 AND E2).  
 RC TISSUE=Brain;  
 RX MEDLINE=89059137; PubMed=3208110;  
 RA Kohlma N., Kato Y., Shiao T., Obata K.;  
 RT "Nucleotide sequences of two embryonic drebrins, developmentally  
 regulated brain proteins, and developmental change in their mRNAs.";  
 RL Brain Res. 464:207-215(1988).

GN [2]



KW	Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;	RL	DNA Res. 3:109-116(1996);
KW	Alternative splicing.	CC	-1 - FUNCTION: PROBABLE METHYLTRANSFERASE.
DOMAIN	1 282	CC	-1 - SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
FT	61 165	CH 1.	
FT	180 282	CH 2.	
FT	REPEAT 309	SPECTRIN 1.	
FT	REPEAT 398	SPECTRIN 2.	
FT	REPEAT 429	SPECTRIN 3.	
FT	REPEAT 533	SPECTRIN 4.	
FT	REPEAT 535	SPECTRIN 5.	
FT	REPEAT 642	SPECTRIN 6.	
FT	REPEAT 644	SPECTRIN 7.	
FT	REPEAT 773	SPECTRIN 8.	
FT	REPEAT 881	SPECTRIN 9.	
FT	REPEAT 1019	SPECTRIN 10.	
FT	REPEAT 1088	SPECTRIN 11.	
FT	REPEAT 1197	SPECTRIN 12.	
FT	REPEAT 1303	SPECTRIN 13.	
FT	REPEAT 1305	SPECTRIN 14.	
FT	REPEAT 1408	SPECTRIN 15.	
FT	REPEAT 1410	SPECTRIN 16.	
FT	REPEAT 1513	SPECTRIN 17.	
FT	REPEAT 1515	SPECTRIN 18.	
FT	REPEAT 1619	PH.	
FT	REPEAT 1725	MISSING ( IN ISOFORM 3 ).	
FT	REPEAT 1727	AVQAAAGLRLGNGYGEAQEAATRLLEK -> MPHYPSCS	
FT	REPEAT 1832	SASLGTPPPQQQOLEARHR ( IN ISOFORM 3 ).	
FT	REPEAT 1834	NOEHLQRQWQWMKLHDLQH -> CLIIHPALLHPWE	
FT	REPEAT 1940	PYLPRSS ( IN ISOFORM 2 ).	
FT	REPEAT 2046	MISSING ( IN ISOFORM 2 ).	
FT	REPEAT 2048	TEKKAKEDSKQPPIPLRKFFDDPTELAKAAKAPLLRRGGY	
FT	REPEAT 2418	E -> PRBDLHPLGPVOPPWQTEKPSLKPKRANKERKAY	
FT	VARSPPLIC 1	RDGTCI ( IN ISOFORM 4 ).	
FT	VARSPPLIC 1258	MISSING ( IN ISOFORM 4 ).	
FT	VARSPPLIC 1258	MISSING ( IN REF. 2 ).	
FT	VARSPPLIC 1287	L -> S ( IN REF. 2 ).	
FT	VARSPPLIC 1310	E -> K ( IN REF. 2 ).	
FT	VARSPPLIC 2113	E -> S ( IN REF. 2 ).	
FT	VARSPPLIC 2154	G -> S ( IN REF. 1 ).	
FT	VARSPPLIC 2155	288982 MW: 52CD87D1D6018CC CRC64;	
FT	CONFLICT 604	Query Match 46.8%; Score 37; DB 1; Length 211;	
FT	CONFLICT 714	Best Local Similarity 50.0%; Pred. No. 2; Mismatches 7; Conservative 7; Matches 7; Gap 1;	
FT	CONFLICT 1189	AC P22734;	
FT	CONFLICT 1193	DT 01-AUG-1991 ( Rel. 19, Created )	
FT	CONFLICT 1331	DT 01-MAY-1992 ( Rel. 22, Last sequence update )	
FT	SEQUENCE 2564 AA:	DT 15-JUN-2002 ( Rel. 41, Last annotation update )	
FT	SEQUENCE 2564 AA:	DE Catechol O-methyltransferase, membrane-bound form ( EC 2.1.1.6 )	
FT	SEQUENCE 2564 AA:	DE (MB-COMT) [ Contains: Catechol O-methyltransferase, soluble form (S-COMT) ].	
FT	SEQUENCE 2564 AA:	DE (S-COMT) .	
FT	SEQUENCE 2564 AA:	DE COMT.	
FT	SEQUENCE 2564 AA:	DE Rattus norvegicus ( Rat ).	
FT	SEQUENCE 2564 AA:	OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
FT	SEQUENCE 2564 AA:	OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
FT	SEQUENCE 2564 AA:	OX NCBI_TAXID=10116;	
FT	SEQUENCE 2564 AA:	OX [1]	
FT	SEQUENCE 2564 AA:	RE	
FT	SEQUENCE 2564 AA:	RE SEQUENCE FROM N.A.	
FT	SEQUENCE 2564 AA:	RE TISSUE=Liver;	
FT	SEQUENCE 2564 AA:	RE MEDLINE=94107221; PubMed=8280056;	
FT	SEQUENCE 2564 AA:	RE Tenhunen J., Ulmanen I.;	
FT	SEQUENCE 2564 AA:	RT "Production of rat soluble and membrane-bound catechol O-	
FT	SEQUENCE 2564 AA:	RT methyltransferase forms from bifunctional mRNAs.";	
FT	SEQUENCE 2564 AA:	RT Biochem. J. 296:595-600(1993).	
FT	SEQUENCE 2564 AA:	RT [2]	
FT	SEQUENCE 2564 AA:	RT SEQUENCE OF 11-264 FROM N.A.	
FT	SEQUENCE 2564 AA:	RT MEDLINE=91033034; PubMed=227437;	
FT	SEQUENCE 2564 AA:	RT Salminen M., Lundstroem K., Tilgmann C., Savolainen R., Kalkkinen N.,	
FT	SEQUENCE 2564 AA:	RT Ulmanen I.;	
FT	SEQUENCE 2564 AA:	RT "Molecular cloning and characterization of rat liver catechol-O-	
FT	SEQUENCE 2564 AA:	RT methyltransferase.";	
FT	SEQUENCE 2564 AA:	RT Gene 93:241-247(1990).	
FT	SEQUENCE 2564 AA:	RT [3]	
FT	SEQUENCE 2564 AA:	RT SEQUENCE OF 1-10 FROM N.A. AND CHARACTERIZATION OF THE TWO FORMS.	
FT	SEQUENCE 2564 AA:	RT MEDLINE=92111472; PubMed=1765063;	
FT	SEQUENCE 2564 AA:	RT Ulmanen I., Lundstroem K.;	
FT	SEQUENCE 2564 AA:	RT "Cell-free synthesis of rat and human catechol O-methyltransferase in vitro.";	
FT	SEQUENCE 2564 AA:	RT Eur. J. Biochem. 202:1013-1020(1991).	
FT	SEQUENCE 2564 AA:	RT X-RAY CRYSTALLOGRAPHY ( 2.0 ANGSTROMS ) OF SOLUBLE FORM.	
FT	SEQUENCE 2564 AA:	RT MEDLINE=94173341; PubMed=8127373;	
RN	[1]	RN	
RN	SEQUENCE FROM N.A.; Pubmed=8905231;	RN	
RN	MEDLINE=97061201;	RN	
RN	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirokawa M., Sugiyama M., Sasamoto S., Kimura T., Hosouchi T., Matsumi A., Nakazaki N., Naoko K., Okumura S., Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;	RN	
RN	"Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";	RN	

Ra Vidgren J., Svensson L.A., Liljas A.; "Crystal structure of catechol O-methyltransferase.";  
 RT  
 RL Nature 368:354-358(1994).  
 CC  
 CC INACTIVATION OF CATECHOLAMINE NEUROTRANSMITTERS AND CATECHOL  
 HORMONES. ALSO SHORTENS THE BIOLOGICAL HALF-LIVES OF CERTAIN  
 NEUROACTIVE DRUGS, LIKE L-DOPA, ALPHA-METHYL DOPA AND  
 ISOPROTERENOL.  
 CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + catechol = S-  
 adenosyl-L-homocysteine + guaiacol.  
 CC -I- COFACTOR: MAGNESIUM.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (isoform S-COMT), Type II  
 membrane protein (isoform MB-COMT).  
 CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; a membrane-bound form/MB-COMT  
 (shown here) and a soluble form/S-COMT; are produced by  
 alternative initiation.  
 CC -I- PTM: The N-terminus is blocked.  
 CC -I- SIMILARITY: TO OTHER MAMMALIAN CATECHOL-O-METHYLTRANSFERASE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: Z12651; CAA78267.1; -; ALT\_INIT.  
 DR EMBL: M60754; AAA40882.1; ALT\_INIT.  
 DR EMBL: M60753; AAA40881.1; ALT\_INIT.  
 DR PIR: JO0787; JO0787.  
 DR PIR: S22050; S22050.  
 DR InterPro: IPR002935; Methyltransf\_3.  
 DR InterPro: IPR00051; SAM\_bind.  
 DR InterPro: IPR00051; Methyltransf\_3.  
 DR PFM: PF01596; Methyltransf\_3; 1.  
 KW Transf erase; Methyltransf erase; Neurotransmitter degradation;  
 KW Catecholamine metabolism; Transmembrane; Signal-anchor; Magnesium;  
 KW Alternative initiation; 3D-structure.  
 FT CHAIN 1 264 CATECHOL-O-METHYLTRANSFERASE, MEMBRANE-  
 FT BOUND\_ISOFORM.  
 FT CHAIN 45 264 CATECHOL-O-METHYLTRANSFERASE, SOLUBLE  
 FT INIT\_MET 44 FOR SOLUBLE ISOFORM.  
 FT TRANSMEM 3 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT ACT\_SITE 187 (POTENTIAL).  
 FT ACT\_SITE 242 184  
 FT METAL 212 184 MAGNESIUM.  
 FT METAL 213 212 MAGNESIUM.  
 FT SEQUENCE 264 AA; 29597 MW; F535DFF49C062854 CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 264;  
 Best Local Similarity 50.0%; Pred. No. 34; Mismatches 6; Conservatve 4; Indels 0; Gaps 0;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MTEQQNRFAGLE 12  
 Db 141. ITQOQMNFAGLO 152  
 RESULT 14  
 ID G3P\_ESCBL STANDARD; PRT; 294 AA.  
 AC P2479; 01-MAR-1992 (Rel. 21, Last created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)  
 DE (Fragment).  
 GN GAP  
 OS Escherichia blattae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC "Crystal structure of catechol O-methyltransferase.";  
 OC Escherichia.  
 OC NCBI\_TaxID=563;  
 RN [1] SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29907, ATCC 33429, and ATCC 33430;  
 RX MEDLINE=92065252; PubMed=1955870;  
 RA Lawrence J.G., Ochman H., Hartl D.L.;  
 RT "Molecular and evolutionary relationships among enteric bacteria.";  
 RL J. Gen. Microbiol. 137:1911-1921(1991).  
 CC -I- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +  
 NAD(+) = 3-phospho-D-glyceraldehyde 3-phosphate + NADH.  
 CC -I- PATHWAY: Second phase of glycolysis; first step.  
 CC -I- SUBUNIT: HOMOTETRAMER.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -I- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
 DEHYDROGENASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: M6358; AAA23837.1; -.  
 DR Pfam: PF00044; gpdh.C; 1.  
 DR EMBL: M6359; AAA23852.1; -.  
 DR PROSITE: PS0071; GPDH; 1.  
 DR Glycolysis; oxidoreductase; NAD.  
 DR HSSP: P06977; IGAD.  
 DR InterPro: IPR000173; GAP\_dhydrogenase.  
 DR Pfam: PF00800; gpdh.C; 1.  
 DR PROSITE: PS0071; GPDH; 1.  
 DR EMBL: M6350; AAA23855.1; -.  
 DR Non\_TER 1  
 FT BINDING 135 135 GLYCERALDEHYDE 3-PHOSPHATE  
 FT ACT\_SITE 162 ACTIVATES THIOL GROUP DURING CATALYSIS.  
 FT NON\_TER 162 294  
 FT SEQUENCE 294 AA; 31372 MW; 18562CB39817788B CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 294;  
 Best Local Similarity 54.5%; Pred. No. 38;  
 Matches 6; Conservatve 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 QRNFGAGLEAA 15  
 Db 69 KRNNEAGVEVA 79  
 RESULT 15  
 ID YD59\_SCPO STANDARD; PRT; 335 AA.  
 AC 010312; Q9TUL8;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6C3.09 in chromosome I.  
 GN SPAC6C3.09.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetidae;  
 OC Schizosaccharomyces.  
 OC NCBI\_TaxID=4896;  
 RN [1] SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vansireels E., Rieder M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Lange I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Burneille B.,  
 RA Goiffau A., Cadieu E., Dreano S., Gloux S., Leilaure V., Mottier S.,  
 RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Talleda V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCormick W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "the genome sequence of *Schizosaccharomyces pombe*";  
 RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 DR EML; 269731; CAB0281.1; -.  
 KW Hypothetical Protein.

SQ SEQUENCE 335 AA; 37948 MW; C1CD4E5850BAD323 CRC64;  
 Query Match 46.8%; Score 37; DB 1; Length 335;  
 Best Local Similarity 62.5%; Pred. No. 43;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 6 WNPAGIEA 13  
 ||| 1:1:  
 Db 290 WNFIGVES 297

Search completed: July 3, 2003, 14:05:51  
 Job time : 7.5 secs

					Gencore version 5.1.6
Copyright (c) 1993 - 2003	Compugen Ltd.				
Run on:	July 3, 2003, 14:03:50 ;	Search time 12.5 Seconds			
		{without alignments}			
Scoring table:	BLOSUM62	115.361 Million cell updates/sec			
Searched:	283224 seqs, 96134422 residues				
Total number of hits satisfying chosen parameters:	283224				
Minimum DB seq length:	0				
Maximum DB seq length:	2000000000				
Post-processing:	Minimum Match 0%				
	Maximum Match 100%				
Listing first 45 summaries					
Database :	PIR73:**				
	1: pir1:*				
	2: pir2:*				
	3: pir3:*				
	4: pir4:*				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.					
		RESULTS			
Result No.	Score	Query	Match Length	DB ID	Description
1	79	100.0	95	A70803	early secretory antigen target esat6 - Mycobacterium tuberculosis (strain H37RV)
2	45	57.0	476	S71360	C;Species: Mycobacterium tuberculosis C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999 C;Accession: A70803; S49174
3	45	57.0	476	H95313	R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
4	43	54.4	217	A82491	; Connor, R.; Davies, R.; Devin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
5	43	54.4	247	H90825	Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Squares, S.
6	42	53.2	369	E86554	Nature 393, 537-544, 1998
7	42	53.2	369	G72069	A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
8	42	53.2	539	E87276	A;Title: Deciphering the genome of Mycobacterium tuberculosis from the complete geno
9	42	53.2	723	E87706	A;Reference number: A70500; MUID:98295987; PMID:9834230
10	41	51.9	176	1 F84864	A;Accession: A70803
11	41	51.9	176	2 H90838	A;Status: nucleic acid sequence not shown; translation not shown
12	41	51.9	176	H85696	A;Molecule type: DNA
13	41	51.9	445	2 D82342	A;Residues: 1-95 <COL>
14	41	51.9	471	2 E869435	A;Cross-references: GB:AL022120; GB:AL123456; NID:93261558; PIDN:CAA17967.1; PDB:e126
15	40	50.6	38	2 H82256	A;Experimental source: strain H37Rv
16	39	49.4	219	2 AG2036	A;Soerensen, A.L.; Nagai, S.; Houen, G.; Andersen, P.; Andersen, A.B.
17	39	49.4	328	2 A70145	submitted to the EMBL Data Library, June 1994
18	39	49.4	393	2 AH2150	A;Reference number: S49174
19	39	49.4	482	2 T01762	A;Accession: S49174
20	39	49.4	714	2 G86844	A;Molecule type: DNA
21	39	49.4	916	2 D83093	A;Cross-references: EMBL:X79562; NID:9531708
22	38.5	48.7	392	2 A87572	C;Genetics:
23	38	48.1	252	2 PC1140	C;Gene: esat6
24	38	48.1	284	2 E82868	Query Match 100.0%; Score 79; DB 2; length 95;
25	38	48.1	297	2 AD2851	Best Local Similarity 100.0%; Pred. No. 1.3e-06; Mismatches 0; Indels 0; Gaps 0;
26	38	48.1	305	1 H75285	Matches 15; Conservative 0; Db
27	38	48.1	319	1 B97628	1 MTEQWNFAGIEAA 15
28	38	48.1	443	2 AE0826	0 MTEQWNFAGIEAA 15
29	48.1				RESULT 2
30					S71360
31					noea protein - Rhizobium meliloti
32					C;Species: Rhizobium meliloti
33					C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 04-Mar-2000
34					C;Accession: S71360; S7158
35					R;Arduel, M.; Lorret, G.; Mallet, F.; Roche, P.; Truchet, G.; Prone, J.C.; Rosenbe
36					submitted to the EMBL Data Library, May 1995
37					A;Reference number: S71360
38					A;Accession: S71360
39					A;Molecule type: DNA
40					A;Residues: 1-476 <ARD>
41					A;Cross-references: EMBL:U26430; NID:91326068; PIDN: AAC44091.1; PID:91326070
42					A;Experimental source: strain RCR2011

R;Ardourel, M.; Lortet, G.; Maillet, F.; Roche, P.; Truchet, G.; Prome, J.C.; Rosenberg, R.; Mol, Microbiol., 17, 667-699, 1995  
 A;Title: In Rhizobium meliloti, the operon associated with the nod box n5 comprises nodL  
 A;Reference number: 871357; MUID:9611489; PMID:8801423  
 A;Accession: S71558  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 304-339 <RAW>  
 A;Cross-references: EMBL:U6430  
 A;Experimental source: strain ncrR201 (=SU47)  
 C;Genetics:  
 A;Gene: noeA  
 C;Superfamily: Rhizobium meliloti noeA protein  
 C;Keywords: modulation  
 C;Keywords: nodulation  
 Query Match 57.0%; Score 45; DB 2; Length 476;  
 Best Local Similarity 63.6%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 5 QWNPAIGEAA 15  
 ::::|::|::|::|  
 Db 100 ENSFAGLKAA 110

RESULT 3  
 H95313  
 Note: host specific nodulation protein [imported] - Sinorhizobium meliloti (strain 1021)  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C;Accession: H95313  
 C;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C. PROC. NATL. ACAD. SCI. U.S.A. 98, 9883-9888, 2001  
 A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti A;Reference number: A95262; MUID:21396509; PMID:11481432  
 A;Accession: H95313  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-476 <KUR>  
 A;Cross-references: GB:AE006469; PIDN:AAK65074\_1; PIDN:g14523508; GSPDB:GN00165  
 A;Experimental source: strain 1021, megaplasmid pSvma  
 R;Galibert, F.; Flan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelauvre, hebuilt; P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: noeA  
 A;Genome: plasmid  
 C;Superfamily: Rhizobium meliloti noeA protein

Query Match 54.4%; Score 43; DB 2; Length 217;  
 Best Local Similarity 75.0%; Pred. No. 5.5; Mismatches 3; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 2 TEQOWNFAGIEA 13  
 ::|::|::|::|  
 Db 170 TEQOWLQAGIHA 181

RESULT 5  
 H90825  
 Hypothetical protein ECS1576 [imported] - Escherichia coli (strain 0157:H7, substrate C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C;Accession: H90825  
 C;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, R.; Ishii, K.; Yokoyama, K.; Han, C gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and 9  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: H90825  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-247 <HAY>  
 A;Cross-references: GB:BA00007; PIDN:BA34999\_1; PIDN:g13361040; GSPDB:GN00154  
 A;Experimental source: strain 0157:H7, substrain RIMD 0509952  
 A;Gene: ECS1576  
 Query Match 54.4%; Score 43; DB 2; Length 247;  
 Best Local Similarity 53.8%; Pred. No. 6.3; Mismatches 3; Indels 0; Gaps 0;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Qy 3 EQWNPAIGEAA 15  
 ::|::|::|::|  
 Db 22 QDNWFGSGIKAA 34

RESULT 6  
 E86554  
 Fe-S oxidoreductase [imported] - Chlamydophila pneumoniae (strain J138)  
 C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C;Accession: E86554  
 R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000  
 A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A;Reference number: A86491; MUID:20330349; PMID:10871362  
 A;Accession: E86554  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-369 <SDQ>  
 A;Cross-references: GB:BA000086; MUID:98978883; PIDN:BA98719\_1; GSPDB:GN00142  
 A;Experimental source: strain J138  
 C;Genetics:  
 A;Gene: CP0513  
 C;Superfamily: hypothetical protein AF0390  
 Query Match 53.2%; Score 42; DB 2; Length 369;

	Best Local Similarity 77.8%; Pred. No. 14; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	Oy	1 MTEQQWNFAGIERA 14
Oy	6 WNPAGIEAA 14	Db	48 LSQHQWDEAGLEAA 61
Db	305 WNWGLGIEAA 313		
	RESULT 7		RESULT 9
G72069	conserved hypothetical protein CP0240 [imported] - Chlamydophila pneumoniae (strains CWI C; Species: Chlamydophila pneumoniae, Chlamydia pneumoniae	E87706	prolyl oligopeptidase family protein [imported] - caulobacter crescentus
N; Alternate names: fes's oxidoreductase	C; Species: Caulobacter crescentus	C; Species: Caulobacter crescentus	C; Species: Caulobacter crescentus
C; Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000	C; Accession: E87706	C; Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001	C; Accession: E87706
C; Accession: G81599	R; Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001	R; Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001	R; Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
R; Kalmann, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999	A; Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.	A; Title: Complete Genome Sequence of Caulobacter crescentus.	A; Title: Complete Genome Sequence of Caulobacter crescentus.
A; Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.	A; Reference number: A87249; MUID:21173698; PMID:11259647	A; Reference number: A87249; MUID:21173698; PMID:11259647	A; Reference number: A87249; MUID:21173698; PMID:11259647
A; Accession: G72069	A; Accession: E87706	A; Accession: E87706	A; Accession: E87706
A; Status: preliminary	A; Status: preliminary	A; Status: preliminary	A; Status: preliminary
A; Molecule type: DNA	A; Molecule type: DNA	A; Molecule type: DNA	A; Molecule type: DNA
A; Residues: 1-369 <R>A>	A; Residues: 1-723 <S>T>O>	A; Residues: 1-723 <S>T>O>	A; Residues: 1-723 <S>T>O>
A; Cross-references: GB:AE001636; GB:AE001363; NID:94376794; PIDN:AAD18653.1; PID:9437680	A; Cross-references: GB:AE005673; NID:913425449; PIDN:AAK25649.1; GSPDB:GN00148	A; Cross-references: GB:AE005673; NID:913425449; PIDN:AAK25649.1; GSPDB:GN00148	A; Cross-references: GB:AE005673; NID:913425449; PIDN:AAK25649.1; GSPDB:GN00148
A; Experimental source: strain CWL029	A; Gene: CC3687	A; Gene: CC3687	A; Gene: CC3687
R; Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, N.; Nucleic Acids Res. 28, 1597-1606, 2000	C; Superfamily: hypothetical protein AF0390	C; Superfamily: hypothetical protein AF0390	C; Superfamily: hypothetical protein AF0390
A; Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.			
A; Reference number: A81500; MUID:20150225; PMID:10684935			
A; Accession: F81599			
A; Status: preliminary			
A; Molecule type: DNA			
A; Residues: 1-369 <R>A>			
A; Cross-references: GB:AE002184; GB:AE002161; NID:97189156; PIDN:AAF38105.1; PID:9718916			
A; Experimental source: strain AR39, HL cells			
A; Genetics: CP0513; CP0340			
C; Genetics: CP0513; CP0340			
C; Superfamily: hypothetical protein AF0390			
Query Match 53.2%; Score 42; DB 2; Length 369;			
Best Local Similarity 77.8%; Pred. No. 14; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
Qy 6 WNWGLGIEAA 14			
Db 305 WNWGLGIEAA 313			
RESULT 8	RESULT 10		
E87276	F64864		
glucose-6-phosphate isomerase [imported] - Caulobacter crescentus	protein-disulfide oxidoreductase (EC 1.8.4.-) dsbB [validated] - Escherichia coli (strain K-12)		
C; Species: Caulobacter crescentus	C; Species: Escherichia coli		
C; Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001	C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002		
C; Accession: E87276	C; Accession: F64864; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Rieblatner, F.R.; Rose, D.J.; Mau, B.; Shao, Y.		
R; Nielsen, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001	A; Title: Identification and characterization of the Escherichia coli gene dsbB, whose product, DsbB, is required for the correct folding of proteins in the periplasm. A; Reference number: A48288; MUID:93348217; PMID:7688471		
A; Title: Complete Genome Sequence of Caulobacter crescentus.	A; Status: preliminary		
A; Reference number: A87249; MUID:21173698; PMID:11259647	A; Reference number: A64720; MUID:97426617; PMID:9278503		
A; Accession: F64864	A; Status: nucleic acid sequence not shown; translation not shown		
A; Molecule type: DNA	A; Molecule type: DNA		
A; Residues: 1-176 <BLAT>	A; Residues: 1-176 <BLAT>		
A; Cross-references: GB:AE00216; GB:U00096; NID:91787417; PIDN:AACT4269.1; PID:917874	A; Cross-references: GB:AE00216; GB:U00096; NID:91787417; PIDN:AACT4269.1; PID:917874		
A; Experimental source: strain K-12, substrain MG1655	R; Missakis, D.; Georgopoulos, C.; Raina, S.		
R; Nielsen, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Proc. Natl. Acad. Sci. U.S.A. 90, 7084-7088, 1993	A; Title: Identification and characterization of the Escherichia coli gene dsbB, whose product, DsbB, is required for the correct folding of proteins in the periplasm. A; Reference number: A48288; MUID:93348217; PMID:7688471		
A; Title: Cloning, sequencing, and expression of the nhab gene, encoding a Na+/H+ anti-A; Reference number: JCI108; MUID:92283803; PMID:1317851	A; Status: preliminary		
A; Reference number: JCI108; MUID:92283803; PMID:1317851	A; Molecule type: DNA		
A; Accession: E87276	A; Residues: MI, 1-176 <MS>		
A; Status: preliminary	A; Cross-references: GB:L03721; NID:9398017; PIDN:AAA23711.1; PID:9398018		
A; Molecule type: DNA	R; Pinner, E.; Padan, E.; Schuhdiner, S.		
A; Residues: 1-39 <STGT>	J. Biol. Chem. 267, 11064-11068, 1992		
A; Cross-references: GB:AE005673; NID:913421349; PIDN:AAK22209.1; GSPDB:GN00148	A; Title: Cloning, sequencing, and expression of the nhab gene, encoding a Na+/H+ anti-A; Reference number: JCI108; MUID:92283803; PMID:1317851		
A; Genetics: CC0222	A; Status: translation not shown		
C; Superfamily: glucose-6-phosphate isomerase	A; Molecule type: DNA		
Query Match 53.2%; Score 42; DB 2; Length 539;	A; Residues: MI, 1-176 <MS>		
Best Local Similarity 50.0%; Pred. No. 22; Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;	A; Cross-references: GB:M83655; NID:92283803; PMID:1317851		
C; Genetics: CC0222	A; Experimental source: strain K-12, substrain WI333		

A:Gene: dsbb; roxb  
A:Map position: 25.5 min  
C:Function:  
A:Description: reoxidizes dsba protein specifically  
A:Note: reaction depends on the presence of oxygen  
C:Superfamily: protein-disulfide oxidoreductase dsbb  
C:Keywords: oxidoreductase; redox-active disulfide; transmembrane protein  
F:16-32/Domain: transmembrane #status predicted <TM>  
F:49-65/Domain: transmembrane #status predicted <TM>  
F:41-44/Disulfide bonds: redox-active #status predicted

Query Match 51.9%; Score 41; DB 1; Length 176;  
Best Local Similarity 60.0%; Pred. No. 9.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EQQNFAGIE 12  
Db 132 ERMDFGLIE 141

RESULT 11

H0838 protein-disulfide oxidoreductase ECS1680 [imported] - Escherichia coli (strain 0157:H7, C:Species: Escherichia coli C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 01-Mar-2002 C:Accession: H90838 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and genomic reference number: H99629; MUID:21156231; PMID:11258796 A:Accession: H90838 A:Status: preliminary A:Molecule type: DNA A:Residues: 1-176 <HAT> A:Cross-references: GB:BA000007; PIDN:BAB5103:1; PID:gi|3361144; GSPDB:GN00154 C:Genetics: C:Gene: ECS1680 C:Superfamily: protein-disulfide oxidoreductase dsbb

Query Match 51.9%; Score 41; DB 2; Length 176;  
Best Local Similarity 60.0%; Pred. No. 9.7;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EQQNFAGIE 12  
Db 132 ERMDFGLIE 141

RESULT 12

H85696 hypothetical protein dsbb [imported] - Escherichia coli (strain 0157:H7, substrain EDL93 C:Species: Escherichia coli C:Accession: H85696 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 03-Jun-2002 R:Perri, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grotbeck, E.J.; Davis, N.W.; Lin, A.; Dimalanta, E.; Potamouidis, K.; Apodaca, Nature 409, 529-533, 2001 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7. A:Reference number: A85480; MUID:21074935; PMID:11206551 A:Status: preliminary A:Molecule type: DNA A:Residues: 1-176 <STOP> A:Cross-references: GB:AE005174; PIDN:gI2514882; PIDN:AGS56036:1; GSPDB:GN00145; UMGP:219 A:Experimental source: strain 0157:H7, substrain EDL933 C:Genetics: C:Gene: dsbb C:Superfamily: protein-disulfide oxidoreductase dsbb

Query Match 51.9%; Score 41; DB 2; Length 176;

RESULT 13

D82342 probable cadaverine/lysine antiborster CadB VC0280 [imported] - vibrio cholerae (strain C:Species: Vibrio cholerae C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001 C:Accession: D82342 R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R. J.; Elmer, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, L.; R.R.; Mekalanos, J.J.; Ventre, J.C.; Fraser, C.M. Nature 406, 477-483, 2000 A:Title: DNA Sequence of both chromosomes of the cholera pathogen vibrio cholerae. A:Reference number: AB2035; MUID:20406833; PMID:10952301 A:Status: preliminary A:Accession: D82342 A:Molecule type: DNA A:Residues: 1-445 <HEI> A:Cross-references: GB:AE004116; GB:AE003852; PIDN:AAF93455:1; GSPDB:GN00154 A:Experimental source: serogroup O1; strain N16961; biotype El Tor C:Genetics: C:Gene: VC0280 A:Map Position: 1 C:Superfamily: L-lysine transport protein

Query Match 51.9%; Score 41; DB 2; Length 445;  
Best Local Similarity 60.0%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 6 WNFAGIEAA 15  
Db 199 WSFGVGVESAA 208

RESULT 14

E69435 hypothetical protein AF1486 - Archaeoglobus fulgidus C:Species: Archaeoglobus fulgidus C:Accession: E69435 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Oct-1999 C:Accession: E69435 R:Klenk, H.-P.; Clayton, R.A.; Tomb, J.-F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod , Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.C.; Gill, S.; Kirkness, E Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Ventre, J.C. Nature 390, 364-370, 1997 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch A:Reference number: A69250; MUID:98049343; PMID:9394975 A:Accession: E69435 A:Status: preliminary; nucleic acid sequence not shown; translation not shown A:Molecule type: DNA A:Residues: 1-471 <KLE> A:Cross-references: GB:AE001000; GB:AE00782; PIDN:AB89769:1; PID:9264 C:Superfamily: Methanobacterium conserved hypothetical protein MTH943

Query Match 51.9%; Score 41; DB 2; Length 471;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EQQNFAGIEAA 15  
Db 45 EQMDFEGKEADA 57

RESULT 15

H92256 hypothetical protein vc0978 [imported] - vibrio cholerae (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*  
C;Date: 18 Aug 2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: H82256  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chadron, D.; Embley, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, E.  
I, R.R.; Mekalanos, J.J.; Ventre, J.C.; Fraser, C.M.  
*Nature* 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A;Reference number: A82035; MVID:20406833; PMID:10952301  
A;Accession: H82256  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-38 <HEI>  
A;Cross references: GB:AE004179; GB:AE003852; NID:9655432; PIDN:AAF94140.1; GSPDB:GN001  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VC0978  
A;Map position: 1

Query Match 50.6%; Score 40; DB 2; Length 38;  
Best Local Similarity 50.0%; pred. No. 2.8;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MTEQQNNFAGIERRA 14  
||| |||: |:  
Db 18 MTLNQWNWIGVGVA 31

Search completed: July 3, 2003, 14:07:18  
Job time : 14.5 secs



LENGTH: 2293  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-368-590-2

Query Match      48.1%; Score 38; DB 4; Length 2293;  
; Best Local Similarity 60.0%; Pred. No. 3.7e+02; Mismatches 2;  
; Matches 6; Conservative 2; Indels 0; Gaps 0;

Qy      5 QWNFAGIKA 14  
Db      719 QWRULSGLRA 728

RESULT 14  
US-08-215-805A-80  
; Sequence 80, Application US/08215805A  
; Patent No. 5559008

GENERAL INFORMATION  
; APPLICANT: Chang, Yung-Fu  
; TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTURELLA  
; TITLE OF INVENTION: SUIS  
; NUMBER OF SEQUENCES: 84

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.3.0

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/215,805A  
; FILING DATE: 22-MAR-1994  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
; NAME: Timian, Susan J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1636  
; TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
; LENGTH: 934 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
; ORGANISM: Pasteurella suis  
; STRAIN: 5943

IMMEDIATE SOURCE:  
; LIBRARY: P. suis DNA in Bacteriophage lambda-dash  
; CLONE: (lambda)yfc33-37

US-08-215-805A-80

Query Match      46.2%; Score 36.5; DB 1; Length 934;  
; Best Local Similarity 50.0%; Pred. No. 2.4e+02; Mismatches 5;  
; Matches 8; Conservative 2; Indels 1; Gaps 1;

Qy      1 MTGQW-----NFGI 11  
Db      463 ITQQQWDNNIGNLAGI 478

RESULT 15  
US-09-149-727-2  
; Sequence 2, Application US/09149727

## ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Virginia C.

REGISTRATION NUMBER: 37,092

REFERENCE/DOCKET NUMBER: 627-196

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-579-667-2

Query Match      Score 48.1%; DB 1; Length 410;  
 Best Local Similarity 41.7%; Pred. No. 52;  
 Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 50 KQRWFGSVRSA 61

RESULT 11

US-08-579-667-6

Sequence 8, Application US/08579667

PATENT NO. 5705624

GENERAL INFORMATION:

APPLICANT: Fitzmaurice, Wayne P.

APPLICANT: Hellmann, Gary M.

APPLICANT: Grill, Laurence K.

APPLICANT: Kumagai, Moto H.

APPLICANT: Dell'a-Cioppa, Guy R.

TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN

TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS

NUMBER OF SEQUENCES: 19

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Virginia C. Bennett

STREET: 1211 East Morehead Street, PO Drawer 34009

CITY: Charlotte

STATE: No. 5705624th Carolina

COUNTRY: USA

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/579,667

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Virginia C.

REGISTRATION NUMBER: 37,092

REFERENCE/DOCKET NUMBER: 627-196

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-579-667-6

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match      Score 48.1%; Score 38; DB 1; Length 410;

Best Local Similarity 41.7%; Pred. No. 52;

Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 50 KQRWFGSVRSA 61

Db 50 KQRWFGSVRSA 61

: : : : :

US-09-830-839-1.ra1

RESULT 12

US-08-579-667-8

Sequence 8, Application US/08579667

PATENT NO. 5705624

GENERAL INFORMATION:

APPLICANT: Fitzmaurice, Wayne P.

APPLICANT: Hellmann, Gary M.

APPLICANT: Grill, Laurence K.

APPLICANT: Kumagai, Moto H.

APPLICANT: Dell'a-Cioppa, Guy R.

TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN

TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Virginia C. Bennett

STREET: 1211 East Morehead Street, PO Drawer 34009

CITY: Charlotte

STATE: No. 5705624th Carolina

COUNTRY: USA

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/579,667

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Bennett, Virginia C.

REGISTRATION NUMBER: 37,092

REFERENCE/DOCKET NUMBER: 627-196

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-579-667-8

Query Match      Score 48.1%; Score 38; DB 1; Length 410;

Best Local Similarity 41.7%; Pred. No. 52;

Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 50 KQRWFGSVRSA 61

RESULT 13

US-09-368-590-2

Sequence 2, Application US/09368590

PATENT NO. 6187563

GENERAL INFORMATION:

APPLICANT: Solimena, Michele

TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR

TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES

FILE REFERENCE: 101918-200 (OCR-41)

CURRENT APPLICATION NUMBER: US/09/368,590

CURRENT FILING DATE: 1998-08-04

EARLIER FILING DATE: 1998-08-07

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 2

Oy 3 EQQWNFGAEAA 14

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-579-667-4

Query Match 53.2%; Score 42; DB 1; Length 413;  
Best Local Similarity 50.0%; Pred. No. 10; 3; Mismatches  
Matches 6; Conservative 3; Indels 0; Gaps 0;

Qy 3 EQWNFNGVKA 14  
Db 50 EQWNFNGVKA 61  
|| ||| :||

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/505,860C

RESULT 8  
US-08-685-808-3  
Sequence 3, Application US/08665808  
Patent No. 6048715

GENERAL INFORMATION:  
APPLICANT: HAINES, CHARLES A., et al  
TITLE OF INVENTION: SEPARATION AND CONCENTRATION SYSTEMS BASED  
ON SOLUBLE OLIGOSACCHARIDE BINDING DOMAINS  
NUMBER OF SEQUENCES: 14  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RAE-VENNER LAW GROUP  
STREET: 260 Sheridan Ave., Ste. 440  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94306

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/505,860C

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rae-Venner, Barbara  
REGISTRATION NUMBER: 32,750  
REFERENCE/DOCKET NUMBER: CBDT.017.00US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 328-4400  
TELEFAX: (650) 328-4477

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 155 amino acids

TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: no

RESULT 9  
US-08-685-808-3  
Sequence 3, Application US/08665808  
Patent No. 6048715

GENERAL INFORMATION:  
APPLICANT: Pitzmaurice, Wayne P.  
APPLICANT: Heilmann, Gary M.  
APPLICANT: Grill, Laurence K.  
APPLICANT: Kumagai, Motoo H.  
APPLICANT: Delila-Cioppa, Guy R.  
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN  
TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS  
NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Virginia C. Bennett  
STREET: 1211 East Morehead Street, PO Drawer 34009  
CITY: Charlotte  
STATE: NC  
COUNTRY: USA  
ZIP: 28234

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/579,667

FILING DATE: 24-JULY-1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rae-Venner, Barbara  
REGISTRATION NUMBER: 32,750  
REFERENCE/DOCKET NUMBER: CBDT.017.00US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 328-4400  
TELEFAX: (650) 328-4477

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 155 amino acids

TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: no

RESULT 9  
US-08-685-808-3  
Sequence 3, Application US/08665808  
Patent No. 6048715

GENERAL INFORMATION:  
APPLICANT: HAYNES, CHARLES A., et al

GENERAL INFORMATION:  
APPLICANT: HAYNES, CHARLES A., et al

TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 99:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 51 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-072-596-99

Query Match 100.0%; Score 79; DB 4; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-07; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTQQQWNFAGIEAA 15  
 Db 1 MTQQQWNFAGIEAA 15

RESULT 5  
 US-08-465-640-2  
 Sequence 2, Application US/08465640  
 PATENT NO.: 5955077  
 GENERAL INFORMATION:  
 APPLICANT: ANDERSEN, Peter  
 APPLICANT: ANDERSEN, Ase Bengaard  
 APPLICANT: HASDORN, Kaare  
 APPLICANT: SORENSEN, Anne Lund  
 TITLE OF INVENTION: TUBERCULOSIS VACCINE  
 NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Browdy and Neimark  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/465,640  
 FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/123,182  
 FILING DATE: 20-SEP-1993

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/DK94/00273  
 FILING DATE: 01-JUL-1994

ATTORNEY/AGENT INFORMATION:  
 NAME: COOPER, IVER P  
 REGISTRATION NUMBER: 28,005  
 REFERENCE/DOCKET NUMBER: ANDERSEN-3A

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 TELEX: 248633

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 95 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-465-640-2

Query Match 100.0%; score 79; DB 2; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-07; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTQQQWNFAGIEAA 15

RESULT 6  
 US-09-001-984C-77  
 Sequence 77, Application US/09001984C  
 PATENT NO.: 6245331  
 GENERAL INFORMATION:  
 APPLICANT: Laail, Suman  
 APPLICANT: Zolla-Pazner, Susan  
 APPLICANT: Beilisle, John T  
 TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE  
 FILE REFERENCE: NYU-011  
 CURRENT APPLICATION NUMBER: US/09/001, 984C  
 CURRENT FILING DATE: 1997-12-31  
 PRIOR APPLICATION NUMBER: 60/034, 003  
 PRIOR FILING DATE: 1996-12-31  
 NUMBER OF SEQ ID NOS: 106  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 77  
 LENGTH: 10  
 TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis strain H37RV

RESULT 7  
 US-08-579-667-4  
 Sequence 4, Application US/08579667  
 PATENT NO.: 5705624  
 GENERAL INFORMATION:  
 APPLICANT: Fitzmaurice, Wayne P.  
 APPLICANT: Hellmann, Gary M.  
 APPLICANT: Grilli, Laurence K.  
 APPLICANT: Kumagai, Monto H.  
 APPLICANT: Della Cioppa, Guy R.  
 APPLICANT: Grilli, Laurence K.  
 TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN  
 TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS  
 NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Virginia C. Bennett  
 STREET: 1211 East Morehead Street, PO Drawer 34009  
 CITY: Charlotte  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 28224  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/579, 667  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bennett, Virginia C.  
 REGISTRATION NUMBER: 37, 092  
 REFERENCE/DOCKET NUMBER: 627-196

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-420-2200  
 TELEFAX: 919-881-1175

INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 413 amino acids

RESULT 2  
US-0818-111-99  
; Sequence 99, Application US/08818111  
; Patent No. 638852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Nerco, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vediwick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF NUMBER OF SEQUENCES: 148 CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP STREET: 6300 Columbia Center, 701 Fifth Avenue CITY: Seattle STATE: Washington COUNTRY: USA ZIP: 98104-7092 COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK COMPUTER: IBM PC compatible OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE: Patentin Release #1.0, Version #1.30 CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056, 556 FILING DATE: 07-APR-1998 CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31, 392 TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4300 TELEFAX: (206) 682-6031 INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 amino acids TYPE: amino acid STRANDEDNESS: single TOPOLOGY: linear  
US-09-056-556-104

Query Match Score 79; DB 4; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 100.0%; Pred. No. 2.7e-07; Mismatches 0; Indels 0; Gaps 0;

RESULT 4  
US-09-072-596-99  
; Sequence 99, Application US/09072596  
; Patent No. 648366  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Nerco, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vediwick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF NUMBER OF SEQUENCES: 350 CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP STREET: 6300 Columbia Center, 701 Fifth Avenue CITY: Seattle STATE: Washington COUNTRY: USA ZIP: 98104-7092 COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK COMPUTER: IBM PC compatible OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE: Patentin Release #1.0, Version #1.30 CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072, 596 FILING DATE: 05-MAY-1998 CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31, 392 TELECOMMUNICATION INFORMATION:  
REFERENCE/DOCKET NUMBER: 210121.417C9

RESULT 3  
US-09-056-556-104  
; Sequence 104, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND NUMBER OF SEQUENCES: 241 CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP STREET: 6300 Columbia Center, 701 Fifth Avenue CITY: Seattle

STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk COMPUTER: IBM PC compatible OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE: Patentin Release #1.0, Version #1.30 CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056, 556 FILING DATE: 07-APR-1998 CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31, 392 TELECOMMUNICATION INFORMATION:  
REFERENCE/DOCKET NUMBER: 210121.417C6

Query Match Score 79; DB 4; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 100.0%; Pred. No. 2.7e-07; Mismatches 0; Indels 0; Gaps 0;

RESULT 1  
QY 1 MTEQQWNFAGIEAA 15  
Db 1 MTEQQWNFAGIEAA 15

On protein - protein search, using sw model

Run on: July 3, 2003, 14:04:30 ; Search time 11 Seconds  
                   (without alignments)  
                   40.122 Million cell updates/sec

Title: US-09-830-839-1

Perfect score: 79

Sequence: 1 MTQQWNFAGIERRA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
                   Maximum Match 100%  
                   Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/1/iaa/5A\_COMB\_pep:\*

2: /cgn2\_6/prodata/1/iaa/5B\_COMB\_pep:\*

3: /cgn2\_6/prodata/1/iaa/6A\_COMB\_pep:\*

4: /cgn2\_6/prodata/1/iaa/6B\_COMB\_pep:\*

5: /cgn2\_6/prodata/1/iaa/PCMS\_COMB\_pep:\*

6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. 1 is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	79	100.0	51 4	US-08-818-112-104
2	79	100.0	51 4	US-08-818-111-99
3	79	100.0	51 4	US-09-056-556-104
4	79	100.0	51 4	US-09-072-556-99
5	79	100.0	95 2	US-08-465-640-2
6	79	100.0	52 10	US-09-001-984C-77
7	42	53.2	413 1	US-08-579-667-4
8	38	48.1	155 3	US-08-865-808-3
9	38	48.1	155 4	US-08-505-880C-3
10	38	48.1	410 1	US-08-579-667-2
11	38	48.1	410 1	US-08-579-667-6
12	38	48.1	410 1	US-08-579-667-8
13	38	48.1	2293 4	US-09-368-590-2
14	36.5	46.2	934 1	US-08-215-805A-80
15	36	45.6	602 4	US-09-149-727-2
16	36	45.6	607 4	US-09-149-727-8
17	36	45.6	608 4	US-08-637-670-36
18	36	45.6	6118 4	US-09-149-727-4
19	36	45.6	3111 2	US-08-460-319-4
20	36	45.6	3111 2	US-08-125-077-4
21	35.5	44.9	593 1	US-08-202-319-12
22	35.5	44.9	593 1	US-08-018-129-5
23	35.5	44.9	593 2	US-08-448-250-5
24	35	44.3	360 4	US-09-509-902A-9
25	35	44.3	600 6	5268463-2
26	35	44.3	602 2	US-08-882-704A-5
27	44.3	602 4	US-09-151-957-5	

## ALIGNMENTS

RESULT 1

US-08-818-112-104 ; Sequence 104, Application US/08818112 ; Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedrick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SPEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

ZIP: 98104-7092

ZIP CODE: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13 MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:

LENGTH: 51 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-112-104

Query Match 100.0%; Score 79; DB 4; Length 51;  
                   Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
                   Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;